**UNIVERSITÉ PARIS OUEST**
**NANTERRE LA DÉFENSE**
*Laboratoire EconomiX – UMR 7235*
*École Doctorale Économie, Organisations et Sociétés – ED 396*

**IFP ÉNERGIES NOUVELLES**
*Direction Économie et Veille*

**CHAIRE ÉCONOMIE DU CLIMAT**
*Axe Transitions Énergétiques*

---

# ÉNERGIE, EROI ET CROISSANCE ÉCONOMIQUE DANS UNE PERSPECTIVE DE LONG TERME

---

*Thèse pour l'obtention du Doctorat de Sciences Économiques*
*de l'Université Paris Ouest Nanterre La Défense*
*présentée et soutenue publiquement le 18 novembre 2016 par*

## Victor Court

**Directeurs de thèse :**

| | |
|---|---|
| *Principal* | M. Pierre-André Jouvet, Professeur des universités en sciences économiques<br>EconomiX, Université Paris Ouest Nanterre La Défense |
| *Adjoint* | M. Frédéric Lantz, Professeur titulaire en économie de l'énergie<br>Centre d'Économie et de Gestion, IFP School |

**Autres membres du jury** :

| | |
|---|---|
| *Présidente* | Mme. Natacha Raffin, Professeur des universités en sciences économiques<br>CREAM, Université de Rouen Normandie |
| *Rapporteur* | M. Patrick Criqui, Directeur de recherche CNRS<br>Laboratoire d'Économie Appliquée de Grenoble, Université Grenoble Alpes |
| *Rapporteur* | M. Pierre-Olivier Pineau, Professeur titulaire en sciences de la décision<br>Département de sciences de la décision, HEC Montréal |
| *Examinateur* | M. Alain Ayong Le Kama, Professeur des universités en sciences économiques<br>EconomiX, Université Paris Ouest Nanterre La Défense |
| *Examinateur* | M. Gaël Giraud, Directeur de recherche CNRS<br>Centre d'Économie de la Sorbonne, Agence Française du Développement |

L'UNIVERSITÉ PARIS OUEST NANTERRE LA DÉFENSE (membre de l'UNIVERSITÉ PARIS LUMIÈRES), IFP ÉNERGIES NOUVELLES et la CHAIRE ÉCONOMIE DU CLIMAT n'entendent donner aucune approbation ni improbation aux opinions émises dans cette thèse: ces opinions doivent être considérées comme propres à leur auteur.

# REMERCIEMENTS

Mes premiers remerciements vont naturellement à Pierre-André Jouvet pour le très grand honneur qu'il m'a fait en acceptant de diriger ma thèse. Il aurait été très difficile pour moi de trouver quelqu'un de plus intéressé, compréhensif et investi que Pierre-André pour diriger mes travaux de recherche. J'ai énormément appris à ses côtés et je considère que nos échanges ont été des moments privilégiés. C'est à Pierre-André que je dois la faisabilité même de cette thèse par l'obtention d'un financement de trois ans au sein du laboratoire EconomiX de l'Université Paris Ouest Nanterre La Défense. Je remercie d'ailleurs Valérie Mignon de m'avoir accueilli au sein de ce laboratoire de renom. Je tiens à exprimer ma gratitude à Frédéric Lantz qui a fait le nécessaire pour que mon projet soit intégré au Département Économie et Veille de l'Institut Français du Pétrole et des Énergies Nouvelles. Outre l'honneur pour moi d'avoir réalisé ma thèse dans cet établissement, je considère que l'encadrement de Frédéric a été précieux. Les rendez-vous que nous avons eus tout au long de ma thèse ont été pour moi aussi importants que l'humeur joviale des déjeuners que nous avons partagés. Durant ces trois années de thèse, j'ai également eu la chance d'être intégré au Pôle Transition Énergétique de la Chaire Économie du Climat et j'adresse donc mes plus sincères remerciements à son fondateur, Christian de Perthuis.

Je remercie madame Natacha Raffin et messieurs Alain Ayong le Kama, Patrick Criqui, Pierre-Olivier Pineau et Gaël Giraud d'avoir accepté de participer à mon jury de thèse. J'espère qu'ils auront trouvé mon travail digne de leur intérêt. Je souhaite adresser un remerciement très particulier à Florian Fizaine dont la contribution à cette thèse est considérable. J'ai eu la chance de rencontrer Florian au tout début de ma thèse alors qu'il était en train de terminer la sienne. Nous nous sommes trouvé de très grandes complémentarités de travail, qui ont abouti à ce jour à deux publications. J'espère sincèrement qu'elles ne constituent que le début d'une longue série de collaborations.

En dehors de l'excellence scientifique des établissements auxquels j'ai été affilié, les compétences et la disponibilité de leur personnel administratif ont également beaucoup compté. Je remercie donc très chaleureusement Béatrice Silva, Véronique Robin, Nasam Zaroualete, Abdou Rabba, et Frédéric Hammerer à EconomiX ; Claude Thirault et Chantal Paron à l'IFP ; et Preety Nadarasapillay à la Chaire Économie du Climat. J'ai rencontré un très grand nombre de personnes qui ont toutes participées à mon enrichissement au sein des institutions que j'ai fréquentées pendant ces trois dernières années. Je me considère assez proche de certaines d'entre elles pour ne mentionner que leurs prénoms ; qu'elles n'y voient aucun manque de respect, bien au contraire. J'adresse donc mes sincères remerciements à : Roger Fouquet, Charles Hall, Jessica Lambert, Carey King, Michael Dale, Claire Carter, François-Xavier Chevallerau, François Diaz Maurin, Jacques Treiner, Christophe Goupil, Mathieu Arnoux, Michel Lepetit, Jean-Marc Jancovici, Cédric Ringenbach, Benoît Thévard, Marc Baudry, Olivier Massol, Benoît Chèze, Emmanuel Hache, Bénédicte, Hugo, Jill, Gabriella, Boris, Raphaël, Jeremy, Yvain, Lara, Magali, Karine, Gilles, Hamza, Olivier, Claire, Damien, Florian (L.), William, Zeynep, Jean-Noël, Fatma, Emmanuel, Antoine, Déborah, Romain, Xavier, Antoine, et Cyril.

Mes co-occupants du bureau 313H d'EconomiX méritent bien sûr des remerciements particuliers. Un grand merci à Stellio pour son entrain et sa bonne humeur infaillible. Sans lui, je n'aurais jamais réussi à activer ma carte IZLY pour bénéficier du tarif réservé au « Personnel UPOND » au restaurant universitaire, ce qui aurait sûrement nuit à ma productivité, ou en tout cas à ma santé financière. Qu'il soit donc remercié pour m'avoir sorti d'un méandre administratif qui en a découragé plus d'un. Stellio mérite un autre remerciement pour le travail qu'il a fourni dans la bonne organisation du TD de Macro-L1 que nous avions en commun. Je pense malheureusement ne pas avoir été à la hauteur de son investissement. Enfin, un grand merci à Lesly et Lisa pour leur gentillesse et leur patience, notamment face à mes jurons à répétition.

S'il est difficile de discerner le moment où un collègue devient un ami, il est en revanche beaucoup plus facile de déterminer à posteriori ceux qui le sont devenus. Clément et Simon font partis de ceux-là pour moi. Face à l'isolement social qu'un travail de thèse peut générer (pour un caractère comme le mien en tout cas), avoir quotidiennement à ses côtés deux collègues aussi inspirants représente un luxe dont j'ai eu la chance de bénéficier. Je les remercie vivement pour tous les échanges que nous avons eus sur des sujets aussi sérieux que l'économie et la politique mais aussi et surtout, sur tous ceux beaucoup plus personnels que nous avons en commun.

Mes deux acolytes de travail n'ont fait que compléter une collection déjà bien fournie de proches qu'il me faut remercier pour leur soutien sans faille tout au long de ces trois années de thèse. Tout d'abord, Boris, Manu et Seb, un grand merci à vous trois pour les séances hebdomadaires d'escalade que j'ai eu le plus grand plaisir à partager avec vous. Je n'aurais jamais réussi à tenir pendant trois ans sans ce parfait moyen de décompression. J'espère que nous aurons encore de belles années de grimpe ensemble.

Ensuite, pour les relations très personnelles que j'entretiens avec chacun d'entre eux et pour toutes les soirées à écumer les *dancefloors* parisiens tous les trois : un grand merci à vous deux Dyen et Batou. Merci aussi pour toutes les pintes et les burgers descendus ensemble à refaire le monde. Ces moments ont largement contribués à mon bien être, et donc à la réussite de cette thèse.

Un grand merci également « aux gars », à savoir à la bande de trublions qui me permettait aussi de sortir de ma thèse et de décompresser, notamment les week-ends et en vacances : Simon, Coco, Thom, ThibSalle, Polo, et Yaniv. Un grand merci à vous les mecs, surtout, s'il-vous-plaît ne changez rien. Continuez à me faire rire comme ça à chaque fois que je vous vois. Un membre éminent de cette bande, dit Dal pour les intimes, a volontairement été oublié de la liste précédente. Il a eu une place particulière dans ma vie ces trois dernières années, celle de colocataire aux côté de mon compère de très longue date, le bien nommé Pocardo. Notre « coloc' Laumière » a été extrêmement importante pour le bon déroulement de ma thèse. Merci à vous deux d'avoir partagé mes quelques moments de doutes, et d'avoir été là tous les jours avec votre bonne humeur pour faire les courses, cuisiner, regarder « *Game of Thrones* », monter un mur en placo, s'enflammer devant les JO et toutes les compétitions sportives que nous avons regardées avec notre système-son-5.1-vidéo-projecteur. J'ai sûrement réussi à rentrer à la maison en râlant de temps en temps, mais vous ne pouvez pas savoir combien de fois vos sourires et vos bonnes blagues m'ont désarmé au point de complètement effacer mon amertume.

La liste suivante énumère des personnes peut être moins proches (notamment géographiquement) mais que j'estime tout autant, et que je souhaite remercier pour cette fameuse question qu'ils m'ont tous posés au moins une fois, « *Alors ta thèse, ça avance ?* » : Chamla, Roxane, Jorge, Walter, François, Vinz, Fred, Guich, Shank, Anaïs, Agathe, Anoul, James, Flora, Louich, Cécile, Yann, Petté, Annab, Gabi, Océ, Tristan, Piddy, Choël, Decam, Monget, Mathilde, Yvon, Lachesnaie, Pierre, Anne-France, Joanna, Élisabeth, et Ophélie.

Enfin, comme de coutumes dans ce genre d'exercice, mes derniers remerciements iront à ma famille. J'ai bénéficié toute ma vie du soutien et de la pleine confiance de mes parents dans tous mes choix. Ils n'ont pas failli à leur exemplarité durant ma thèse et je les en remercie sincèrement. Un grand merci aussi à ma sœur Céline, à mes deux frères Germain et Charles, à ma belle-sœur Isabelle, à mes neveux et nièces Louise, Pierre, Oscar et Lucie, à mes quatre grands-parents Nani, Nano, Mamone et Micha, à mes oncles et tantes Didier, Lylian, Philippe, Jean-Jacques, Laurent, Jean-Michel et Michelle, à ma cousine Keelin et à mes deux cousins Hugo et Paul. Durant ces trois dernières années, ces proches ont toujours su me communiquer leur soutien et je les en remercie chaleureusement.

# PUBLICATIONS

The work presented in this thesis has been valorized through different publications. As can be seen below, in addition to a common work with my two thesis directors Pierre-André Jouvet and Frédéric Lantz, my work has also involved several close collaborations with Florian Fizaine. Hence, I choose to sometimes keep plural pronouns like *we* in this thesis as its content is not a pure individual achievement but the result of fruitful partnerships.

*International Peer-Reviewed Journals:*

- Fizaine, F., Court, V., 2015. Renewable electricity producing technologies and metal depletion: a sensitivity analysis using the EROI. *Ecological Economics*, 110, pp.106–118.

- Fizaine, F., Court, V., 2016. Energy expenditure, economic growth, and the minimum EROI of society. *Energy Policy*, 95, pp.172–186.

- Court, V., Fizaine, F., 2016. Estimations of the global energy-return-on-investment (EROI) of coal, oil, and gas in the very long-term. *Ecological Economics*, under revision.

- Court, V., Jouvet, P.A., Lantz, F., 2016. Long-term endogenous economic growth and energy transitions. *Energy Journal*, under revision.

*Non Peer-Reviewed Journals:*

- Court, V., 2015. Énergie nette et EROI (Energy-Return-On-Investment) : Une approche biophysique de la transition énergétique. Série Informations et Débats de la Chaire Économie du Climat, n°40.

# ENERGY, EROI, AND ECONOMIC GROWTH IN A LONG-TERM PERSPECTIVE

# TABLE OF CONTENTS

# INTRODUCTION

> *"It ain't what you don't know that gets you into trouble.*
> *It's what you know for sure that just ain't so."*
>
> Mark Twain

## MOTIVATION

In the early twenty-first century, perhaps more than ever before, people are inquisitive about how wealth is created in a given society, what causes it to rise and fall, and how it is distributed among people. Set against the scale of human history as a whole, these questions are in fact quite recent. For centuries, theologians and political philosophers were attracted rather by reflections about whether social harmony and individual freedom could be compatible. It is only in the last two hundred years or so that *economics* has come to be identified as a distinctive scientific field, the purpose of which is to study how the natural propensity of people to truck, barter, and exchange (Smith 1776) manifests itself at the aggregate level.

What people truck, barter and exchange are goods and services that ultimately come from the transformation of raw materials from nature. By convention, value added is positively accounted for during the successive processing of intermediate materials into final goods and services, and the aggregation of all value added represents the total output of the economy. ***Economic growth*** formally represents the annual rate of growth of the macroeconomic output (generally the real gross domestic product or real GDP expressed in constant currency) and the question of its origin and evolution remains the deepest mystery of economics.

Many people, consciously or not, seem to consider that economists alone can legitimately try to answer these two fundamental questions: Where does economic growth come from? Why are some countries much richer than others? This attitude toward partitioning complex problems is really recurrent in modern societies. Because questions become increasingly specific and require more detailed knowledge than ever, science has been divided into multiple disconnected areas that all try to carry out precise investigations. But multipolar approaches are needed in order to respond intelligently to such complex societal questions. The aim of this thesis is to give more space to the diverse interactions that innately exist between history, economics, natural sciences, mathematics, and physics, in order to address the essential question of the origin of economic growth.

The basic representation of human society that economists have gradually conceptualized is highly surprising: it is a productive system creating goods and services using people's labor and the past accumulation of physical and human capital, without the clear necessity to use raw materials. For most mainstream economists, the possibility of increasing economic output is primarily defined by the combination of available factors of production (labor, physical and human capital) with *technological change*.[1] Economists have difficulties in precisely defining what technological change is and explaining where it comes from, but it is surely an essential component of the economic process. Furthermore, since technological change has always been increasing in the past, mainstream economists generally assert that there is no reason to think that it will not continue to do so.[2]

But even mainstream economists recognize that technological progress and the accumulation of production factors are only *proximate causes* of growth. These factors direct the economic growth process in the short run, but they cannot explain why approximately two hundred years ago some privileged regions (Western Europe and North America) underwent an Industrial Revolution that launched them on a path of relatively sustained high growth rates compared to the previous millennia during which all regions of the world were trapped in a state of Malthusian near-stagnation. One of the most important questions, if not *the* most important question in economics is to understand why these precise regions of the world escaped from Malthusian stagnation at an early date through industrialization, whereas others have had a delayed take-off and seem to be catching up more or less rapidly (respectively Eastern Asia, South and Central America on the one hand, Africa and South Asia on the other). To explain this phenomenon of *Great Divergence*, some economists and historians think that so-called *ultimate*, but more properly *deep-rooted*, *causes* of a biogeographical, cultural, institutional, or accidental nature must be considered if we are to understand the process of long-term economic growth. Of course, scholars do not agree on the relative importance of these factors in explaining why some countries are so rich and others so poor. Endless debates have preoccupied economists/historians on this subject for the last two hundred years, and there have been attempts for some years now to finally build a *Unified Growth Theory* (UGT). As shown in this thesis, a really important aspect of the economic growth process has been forgotten, and that until acknowledged, every UGT will be flawed.

In really simple terms, a real human society is a productive system transforming natural resources into goods and services that people require to satisfy a given standard of living. And among all these natural resources that are mostly forgotten in mainstream economic theories, one in particular seems obviously primordial if the human productive system is to function: ***energy***. The first part of this thesis will show that the only possible ultimate or fundamental cause of growth is energy consumption, or more precisely ***useful exergy consumption***, which is the quantity of primary energy extracted from the environment, transformed into usable forms and dissipated by the economic process. The role of energy (more accurately exergy) must be properly understood to form the basis of a coherent UGT. Building such a comprehensive UGT

---

[1] Throughout this thesis the terms *technological change* and *technological progress* are used interchangeably. I see no apparent formal differences between the two but I usually prefer technological change. The reason is simply that for me technological *progress* contains a normative dimension in the sense that every technological modification would necessarily be better for the economic system and, *a fortiori*, for people's welfare, which I think is absolutely not true.

[2] The leading economic growth expert Philippe Aghion supports this idea in his numerous peer-reviewed articles and in this short French video: http://www.francetvinfo.fr/economie/croissance/video-est-ce-la-fin-de-la-croissance_809791.html.

is a long-term research goal and the present thesis should be understood as the first step along that path. For now, it is important to detail the research question investigated in this thesis and what it tries to accomplish.

## RESEARCH QUESTION

If the role of energy consumption as the fundamental cause of economic growth is not well recognized among scholars, the addiction of modern economies to fossil energy resources has been clear in the public sphere since the two oil crises of the 1970s. Considering that fossil energy resources exist as stocks and are therefore ultimately limited in amount, considerable emphasis is now placed on the increasing need to use renewable energy forms. But it must be highlighted that a renewable energy resource is also constrained, not in terms of total recoverable quantity, but in the quantity available for a given time and in particular for a year. This question of the magnitude of a nonrenewable stock or a renewable flow of energy can be understood as the *availability* of energy resources. The different questions regarding the availability of energy resources and the supply mix that a country should choose have monopolized almost all economists and policy makers' attention regarding the potential of energy as an economic constraint. But they have largely ignored another major concept: ***net energy***. Net energy is basically calculated as the gross energy produced minus the energy invested to obtain that energy. A derived idea of net energy is the ***energy-return-on-investment***, abbreviated to ***EROI***, which characterizes, not the availability, but the *accessibility* of an energy resource. The EROI of an energy resource defines the amount of energy that must be invested to exploit a given energy resource. It represents the difficulty of extracting *primary energy* from the environment and delivering it to society. But primary energy *per se* (coal, oil, gas, solar radiant energy, etc.) is of little use for the economic process as it must be converted into *final energy* forms (refined liquids, gas, heat, electricity) that are ultimately dissipated to provide *useful energy* services (light, heat, and motion). Moreover, across this array of energy forms (primary, final, useful), the fundamental laws of thermodynamics stipulate that only a part of the energy, which is called ***exergy***, is productive in the physical sense and hence can be used up in the economic process.

These notions have been conceptualized by pioneering thinkers in ecology and physics and applied to many different systems, including the economy. Yet, mainstream economists have ignored the importance of these concepts in their theories. The questions that immediately come to mind are: How can theories that disregard the physical essence of the economic system explain past historical growth? Why does mainstream economics ignore the importance of natural resources, and more specifically energy, in the growth process? Are energy availability and accessibility really important control variables of the economic growth process? Is it possible to equate the aggregated technological progress of the economy with its primary-to-useful exergy conversion efficiency? To summarize, the fundamental research question of the present thesis is:

***What is the importance of energy for economic growth?***

## ORGANIZATION OF THE THESIS

This dissertation is composed of seven chapters over which two separate but related issues are explored. Chapters 1 to 3 form an original essay assessing whether net energy (exergy) has played a major role in the economic growth process so far, while Chapters 5 to 7 correspond to published papers investigating what the net energy (exergy) constraint could imply in a future where societies will ultimately have to make a complete transition towards renewable energy. With its central position, Chapter 4 provides some answers to both issues. This hybrid form of organization can be detailed as follows.

Chapter 1 first describes what can be considered as the four main facts of (very) long-term economic growth (transition from stagnation to growth, the Great Divergence, the interdependence of energy consumption and technological change, and the hierarchized-nested adaptive cycle dynamics). Then, the different so-called ultimate causes of growth (biogeography, culture, institutions, and luck), that should preferably be termed deep-rooted factors, are presented in detail.

Chapter 2 pursues the review of the existing theories of growth that focus on the modern regime and its proximate causes, such as technological progress and the accumulation of physical and human capital. The recent attempts to build a Unified Growth Theory are also analyzed in this chapter. This extensive review of all mainstream economic growth theories is indispensable to show that they fail to properly explain the growth process in the long-term.

Chapter 3 starts by analyzing the misguided reasons for overlooking natural resources, and in particular energy, in mainstream economics. The fundamental laws of thermodynamics and essential concepts such as exergy and entropy are then described. This biophysical/thermodynamic approach is essential to a proper understanding of crucial role of energy in the economic system, and of the point that only useful exergy consumption can be considered as the fundamental cause of economic growth.

Chapter 4 presents in detail the static (meaning for a given representative year) calculation methodology of the EROI of a given energy system, the different controversies surrounding such calculation, and hence the limits of this concept. Then, representative EROIs of past and current energy systems are presented. Since so far past estimations of fossil fuel EROIs have been scarce and highly speculative, a price-based methodology is presented to estimate the long-term evolution of coal, oil, and gas EROI. These results are compared to a new theoretical dynamic function of the EROI. Finally, the implications of the probable future decrease in the societal EROI are discussed in the context of the energy transition from fossil to renewable energy.

Chapter 5 presents another important future implication of the energy transition: the degradation of the EROI of renewable technologies due to the strong interrelation of the energy and metal sectors. A methodology to assess the sensitivity of the EROI of a given energy system to the qualitative depletion (i.e. ore grade degradation) of its constient metals is elaborated. Simulations are then performed with up-to-date data.

Chapter 6 investigates the relation between energy expenditure, economic growth, and the minimum required EROI of society. Energy expenditure is estimated for the USA and the world from 1850 to 2012, and for the UK from 1300 to 2008. Concentrating on the USA, energy expenditure estimates are used to show that, statistically, the US economy cannot afford to spend more than 11% of its GDP on energy if positive economic growth is desired. Given the current energy intensity of the US economy, this translates into a minimum EROI of approximately 11:1 (or a maximum tolerable average price of energy of twice the current level). Furthermore, Granger econometric tests consistently reveal a one way causality running from the level of energy expenditure (as a fraction of GDP) to economic growth in the USA between 1960 and 2010.

Chapter 7 is intended to build a bridge between the endogenous economic growth theory, the biophysical economics perspective, and the past and future transitions between renewable and nonrenewable energy forms that economies have had and will have to accomplish. The model supports the evidence that historical productions of renewable and nonrenewable energy have greatly influenced past economic growth. Indeed, from an initial almost-renewable-only supply regime, the model reproduces the increasing reliance on nonrenewable energy that has allowed the global economy to leave the state of economic near-stagnation that had characterized the largest part of its history. Simulations help define the circumstances for which the inevitable future transition towards complete renewable energy could have negative impacts on economic growth (peak followed by degrowth phase).

The conclusion can be analyzed in two parts. The first part of this thesis (Chapters 1 to 3) highlights the adequacy of the biophysical/thermo economics approach for understanding the economic growth phenomenon. If conventional theories are unable to correctly explain the four long-term facts of economic growth (transition from stagnation to growth, Great Divergence, energy consumption–technological change interdependence, and hierarchized-nested adaptive cycles dynamics), this thesis shows that, at least for the first three of them, the role played by energy in the economic system is primordial. Higher energy availability and accessibility are predominant in explaining that the onset of the Industrial Revolution occurred in Britain and not elsewhere. The local energy availability and accessibility, and the magnitude and time differences in the spread of technologies that enable an increase in the aggregate primary-to-useful exergy conversion efficiency have largely defined the direction of the Great Divergence. As a consequence, the future economic growth of countries will depend essentially on (i) the continued increase in the aggregate efficiency of primary-to-useful exergy conversion, and/or (ii) the continued increase in the extraction of available primary exergy resources. The former point has already been discussed but the later needs to be addressed in terms of net energy (exergy).

The second part of this thesis (Chapters 4 to 7) suggests that maintaining a future high net energy supply is likely to become increasingly difficult given the past evolution of fossil fuel EROIs and considering the current low EROIs of renewable energy-producing technologies towards which industrial societies are supposed to make a transition. There are of course significant opportunities for maintaining a high societal EROI or adapting to decreasing EROIs. But from a systemic point of view, industrialized societies seem not to be designed to run with low-density energy resources that come with low EROIs. Until proven otherwise, high economic growth is only possible if high-density energy resources infuse the economic system and allow physical and human capital accumulation, the establishment of inclusive institutions, higher material standards of living, higher qualitative leisure, and in summary greater welfare for people.

# CHAPTER 1

# FACTS AND DEEP-ROOTED CAUSES OF ECONOMIC GROWTH

> *"Perhaps, the destiny of man is to have a short, but fiery, exciting and extravagant life rather than a long, uneventful and vegetative existence."*
>
> Nicholas Georgescu-Roegen

The purpose of this first chapter is to anchor the thesis in historical facts and existing economic theories of long-term economic growth. Many facts describe economic growth, I will present here what I think are the four most important ones: the global dynamics from stagnation to growth over the last millennia, the Great Divergence across regions of the world during the last two to three centuries, the presence of embedded cycles in economic history, and the close relation between major technological changes and energy consumption. I will then review the different narrative-based theories advanced by historians to explain these long-terms patterns. Those approaches can be classified according to the nature of the deep-rooted causes they rely on: geography, culture, institutions, and luck.

## 1.1 SUCCESSIVE STATES, GREAT DIVERGENCE, AND ADAPTIVE CYCLES

### *1.1.1 Successive States: Malthusian Epoch, Post-Malthusian Regime, and Modern Growth Era*

#### The Malthusian Epoch

The most remarkable transformation in the story of mankind is surely the transition from the epoch of slow technological progress and economic near-stagnation that has characterized most of history to the recent state of sustained economic and technological growth. For millennia before the Industrial Revolution of the mid-eighteenth century, living standards in the world economy were almost stagnant. In this *Malthusian Epoch*, population and per capita income remained stable in the absence of technological change or increasing land availability. By contrast, temporary gains in income per capita were generated during periods of technological improvement or increasing land availability, but they were quickly offset by population increases (Galor 2011, p.9). Paradoxically, in the Malthusian Epoch, sudden falls in population caused by wars or epidemics induced momentary increases in income per capita which then vanished due to population catch-ups. Hence, income per capita fluctuated around an almost flat (i.e. slightly increasing) trend in the *Malthusian Trap*. This process characterized the existence of *Homo sapiens* hunter-gatherers from the time they first appeared in Africa roughly 150,000 years ago to their exodus from the East African Rift approximately 60,000 years ago,[1] and up to the Neolithic Revolution that triggered the onset of agricultural communities about 10,000 years ago in various places (Hilly Flanks, Yellow and Yangtze Valleys, the Indus Valley, Mexico, Peru, and perhaps the Eastern Sahara and New Guinea). This Malthusian Trap was still active in the subsequent emergence of cities, states, and nations, until the eve of the Industrial Revolution. Figure 1.1 shows the change in the average global wealth per capita, or more formally the average gross world product (GWP) per capita in 1990 International Geary–Khamis dollars,[2] from 150,000 BCE to 2010 CE.[3] Of course, estimates of GWP per capita do not make real sense before 1950 CE (when international accounting rules were established), and considerable debate still goes on about the proper way to compare material standard of living among countries and over time (Prados de la Escosura 2015). Nevertheless, with all due caution regarding estimates of GWP per capita over such a long time frame, Figure 1.1 shows how much the modern condition that millions of people enjoy nowadays is a *singularity* in the broad human adventure.

Contrary to what Figure 1.1 might suggest, the Malthusian Epoch must not be understood as a completely static regime. Without any doubt technology improved considerably between the Stone Age and the dawn of the Industrial Revolution (see Section 1.2). Yet, it is important to stress that during the Malthusian Epoch, any increase in food or manufactured goods production generated by technological progress or land expansion was primarily channeled toward an increase in the size of the population, providing only a tenuous increase in income per capita. This is because in the Malthusian Epoch, technological progress allowed first of all an increase in the carrying capacity of the environment, i.e. the maximum number of people that the environment could sustain *ad infinitum*. That is why, intensifying modes of food production associated with increasing societal complexity provided support for increasing

---

[1] This phenomenon, known as the "Out of Africa theory" (OOA), the "recent single-origin hypothesis" (RSOH), or "recent African origin" model (RAO), is the most widely accepted model of the geographic origin and early migration of anatomically modern humans. The theory argues that *Homo sapiens*, which is thought to have emerged roughly 150,000 years ago from a common *hominid* ancestor, left Africa approximately 60,000 years ago in a single wave of migration which populated the world, completely replacing older human species such as *Homo erectus* and *Homo neanderthalensis* (Stringer 2003; Liu et al. 2006). An alternative theory is the "multiregional evolution accompanied by gene flow" hypothesis, according to which different evolving human populations continually divided and reticulated and hence allowed both species-wide evolutionary change and local distinctions and differentiation. Contrary to its detractors' assertions, the "Multiregional Hypothesis" does not imply independent multiple origins or simultaneous appearance of adaptive characters in different regions with subsequent parallel evolution (Wolpoff et al. 2000).

[2] The 1990 International Geary-Khamis dollar (Int. G-K. $1990, or more simply $1990), more commonly known as the international dollar, is a standardized and fictive unit of currency that has the same purchasing power parity as the U.S. dollar had in the United States in 1990.

[3] CE is an abbreviation for "Common/Current Era" and BCE is an abbreviation for "before the Common (or Current) Era". The CE/BCE designation uses the same numeric values as the traditional Anno Domini (AD) year-numbering system introduced by the sixth-century Christian monk Dionysius Exiguus, intending the beginning of the life of Jesus to be the reference date (hence dates before are labeled "Before Christ", i.e. BC). The two notations, CE/BCE and AD/BC, are numerically equivalent and neither includes a year zero. Thus "2010 CE" corresponds to "AD 2010", and "10,000 BCE" corresponds to "10,000 BC".

population densities, from hunting-gathering, pastoralism, shifting farming, and traditional farming to modern farming (Figure 1.2).

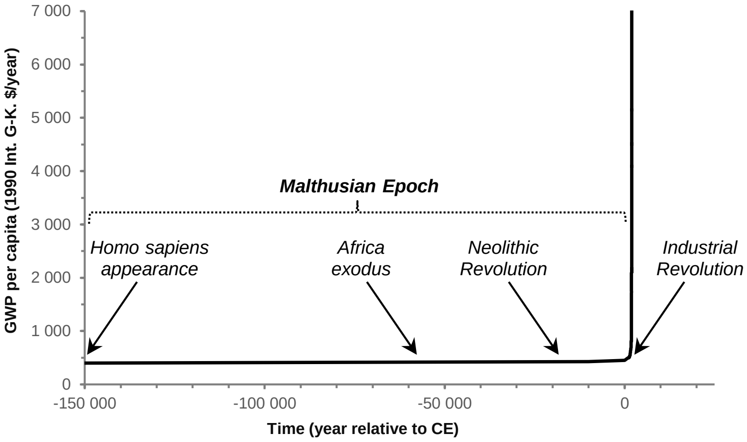


**Figure 1.1 Change in the average GWP per capita, 150,000 BCE–2010 CE.**
**Data source: Maddison (2007) for 1–1820 CE, The Maddison Project (2013) for 1820–2010 CE. Before 1 CE, educated guesses based on Maddison (2007) have been put at $400/yr in 150,000 BCE, $425/yr in 10,000 BCE, $440/yr in 5000 BCE, and $450/yr in 1 CE.**

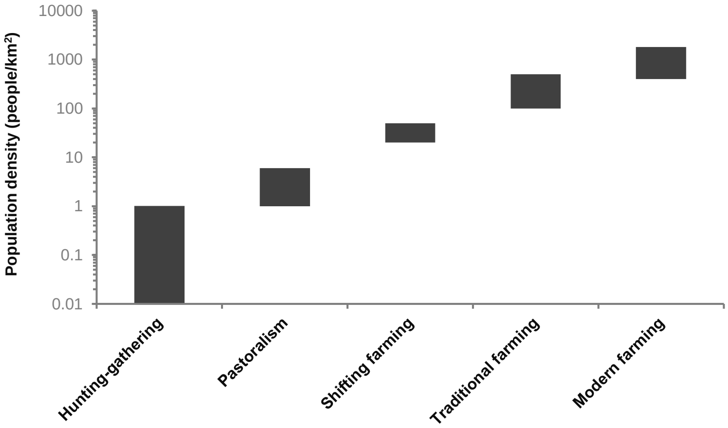


**Figure 1.2 Population densities sustainable by intensifying modes of food production.**
**Source: reproduced from Smil (2008, p.149).**

The first consequence of this fact is that the global population increased from about 2–4 million in 10,000 BCE to 750–800 million just before the Industrial Revolution. The second consequence is that, in the Malthusian Epoch, cross-county differences in technology and land productivity translated into distinct population densities, but impacts on standards of living were merely transitory (Boserup 1965; Kremer 1993; Ashraf & Galor 2011). In the Malthusian state, improving technology led to an increase in the carrying capacity of the local environment, and hence to greater population densities but no higher standards of living, whereas loss of technological leadership or political disruption brought about a decrease of the carrying capacity of the local environment, and hence population density declined which could generate a temporary fall in the standard of living (see Figure 1.3 for a clear characterization of these mechanisms in Egypt from 4000 to 150 BCE).

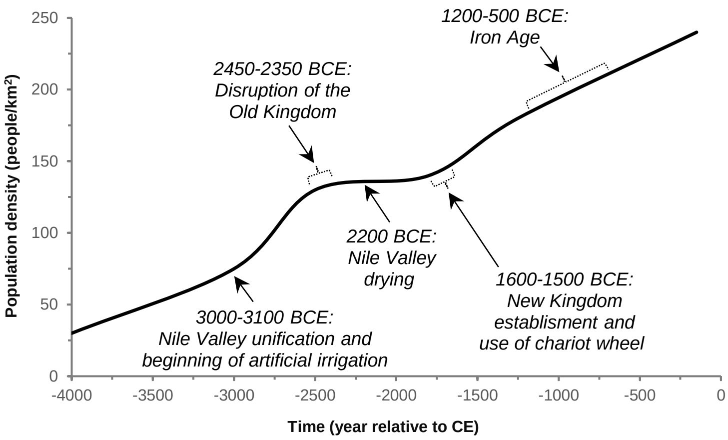


**Figure 1.3 Population density in Egypt, 4000–150 BCE.**
**Data source on population density: Butzer (1976, p.83). Historical annotations: Morris (2010).**

Additionally, there is no doubt that during the Malthusian Epoch, the level of economic and political organization of societies (i.e. societal complexity) increased gradually, but neither steadily nor equally. Technological and political leadership never remained in the same place throughout those years. The Western and Eastern cores alternatively shared the leading role of development in successive empires. Moreover, throughout these millennia the different economies of the world continually increased their exchanges of goods, knowledge, cultural habits, people, and diseases (Morris 2010). Yet, the Malthusian Trap was indeed an epoch of stagnation since historical data indicate that life expectancy, fertility rates, calorie consumption, and stature did not differ significantly between hunter-gatherers of the Mesolithic, cultivator-pastoralists of the Neolithic, and farm laborers and craftsmen of the Middle Ages (Clark 2007).

### Post-Malthusian Regime and Modern Growth Era

In contrast, around 1825 economic production had intensified relative to population in enough countries for mankind to enter a *Post-Malthusian Regime* at the global scale (Figure 1.4). During the Post-Malthusian Regime, the average growth rate of output per capita in the world increased from 0.05% per year in 1500–1820 to 0.5% per year in 1820–1870, and 1.3% per year in 1870–1913 (Maddison 2007, p.383). At first, the rapid increase in income per capita in the Post-Malthusian Regime induced a rise in population. The growth rate of the world population increased from an average of 0.27% per year in 1500–1820 to 0.4% per year in 1820–1870, and to 0.8% per year in 1870–1913 (Maddison 2007, p.377). Hence, evidence suggests that the Malthusian mechanism linking increasing income to increasing population growth remained active at first in the Post-Malthusian Regime. But at some point in the early process of industrialization, the level of capital accumulation and technological progress allowed income per capita to rise by counteracting the effect that higher population growth had on diluting the economic product. In the Post-Malthusian Regime, the fertility rate, birth rate, and death rate all declined compared to the Malthusian Epoch, whereas life expectancy, literacy rate, industrialization and urbanization levels increased. Ultimately, most technologically leading regions experienced a demographic transition (Figure 1.5) which further increased the growth rates of income per capita compared to population growth rates. World output per capita grew by 0.8% per year in 1913–1950 and 2.9% per year in 1950–1973, population grew by 0.9% and 1.9% per year over the same periods respectively (Maddison 2007, pp.377–383). This ended the Post-Malthusian Regime and opened the *Modern Growth Era* of sustained high economic growth around 1950 at global scale (Figure 1.4). In the Modern Growth Era (which, of course, was attained before 1950 by leading countries such as the USA, the UK, France, and Germany), population growth no longer offsets the rise in aggregate income that is enabled by ever increasingly efficient use of accumulating production factors.

Understanding the causes of the escape from the Malthusian Epoch is the first prerogative that any good theory of economic growth should accomplish. More precisely why did mankind remain in the Malthusian Epoch for so many millennia? Why did the escape from the Malthusian Trap only occur two centuries ago? Why not earlier or later? Was this pattern preordained by some long-term lock-in factors? Or is the explanation to be sought in the salutary combination of particular, and possibly contingent,[1] events of the eighteenth century?

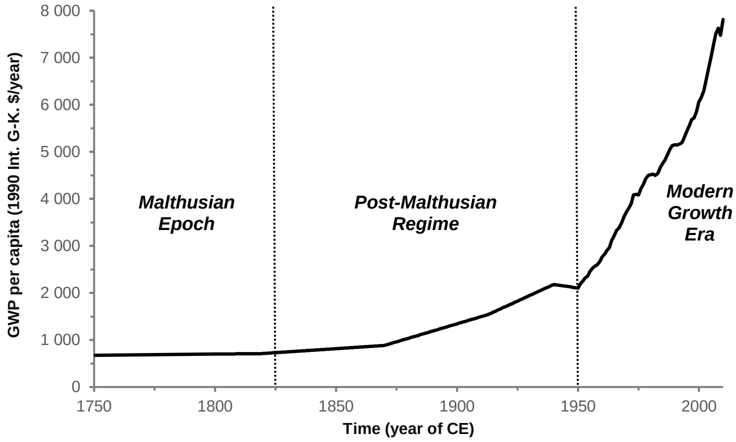


**Figure 1.4 Evolution of the average GWP per capita, 1750–2010 CE. Data source: The Maddison Project (2013).**

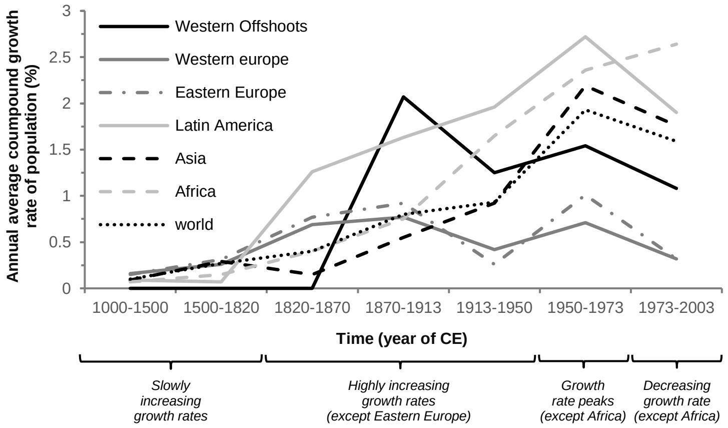


**Figure 1.5 The demographic transition across regions and the world. Data Source: Maddison (2007, p.377).**

---

[1] The occurrence of a contingent event is subject to chance, it can occur or exist only if certain circumstances are present.

### *1.1.2 Great Divergence across Regions of the World*

The global dynamics of Figure 1.4 hide an important reality: the timing and the magnitude of the take-off from the Malthusian Epoch differed among regions of the world. The initial take-off of England from the Malthusian Epoch was associated with the Industrial Revolution that started there in 1750–60, and then spread to Western Europe and the Western Offshoots (USA, Canada, Australia, and New-Zealand) during the first part of the nineteenth century. The take-off of Latin America and West Asia took place toward the beginning of the twentieth century, whereas East Asia and Africa's take-offs were further delayed well into the twentieth century (Galor 2011, p.18). The differential timing of the take-off from stagnation to growth among regions of the world and the associated variations in the timing of their demographic transitions led to the phenomenon called the *Great Divergence*.

As shown in Figure 1.6, some regions (Western Offshoots, Western Europe) have excelled in the growth of income per capita,[1] others have dominated in population growth and hence low income per capita (Eastern Asia, Africa), while a third group has followed the evolution of the global average income per capita (Western Asia, Latin America). More precisely, the GDP per capita spread between the richest and the poorest regions (Western Offshoots and Africa respectively) has widened considerably from a modest ratio of 3:1 in 1800 to an impressive 18:1 in 2000. Of course, inequality *within* societies was already higher in British or Chinese proto-industrial economies (i.e. economies with handicraft manufacturing for the market rather than for home use) of the eighteenth century compared to super-egalitarian bands of Paleolithic hunter-gatherers. And even if historical evidence (Pomeranz 2000; Morris 2010) suggests that momentary technological hegemonies were possible (such as the early but important use of coal and coke for iron smelting around the Yellow delta of China in the eleventh century), inequalities *across* regions have never been as marked as nowadays.

It is one thing to have an initial Great Divergence across regions in the early phases of industrialization two hundred years ago, and quite another thing to see a persistence of regional and national inequalities. Global inequality measured by average national income per capita has increased continuously during the last two hundred years, seems to have peaked around 2000, and remained stable since (Milanovic 2011; 2012). Moreover, national average comparison is a one concept for assessing global inequality, taking into account within-country inequalities is another. In this case, world distribution of income worsened from the early nineteenth century up to World War II and after that seems to have stabilized or to have grown more slowly (Bourguignon & Morrisson 2002), up to the point that global inequalities among citizens of the world appear to have been stable in the last decade (Milanovic 2012). *Within-country* wealth distribution has become (on average) more egalitarian from 1800 to the present, but it has not offset the increasing *between-country* inequality because the relative weight of the latter in the total global inequality of citizens has been increasing over the same period (Bourguignon & Morrisson 2002; Milanovic 2011).

---

[1] As presented in Chapter 2 and visible in Figure 1.6, the most developed economies (Western Offshoots and Europe) are said to be on a *balanced growth path* (BGP) since World War II (WWII). According to Acemoglu (2009, p.57), "balanced growth refers to an allocation where output grows at a constant rate and capital-output ratio, the interest rate, and factor shares remain constant". Mainstream economic models focus on the conditional existence of such state and its possible impediments.

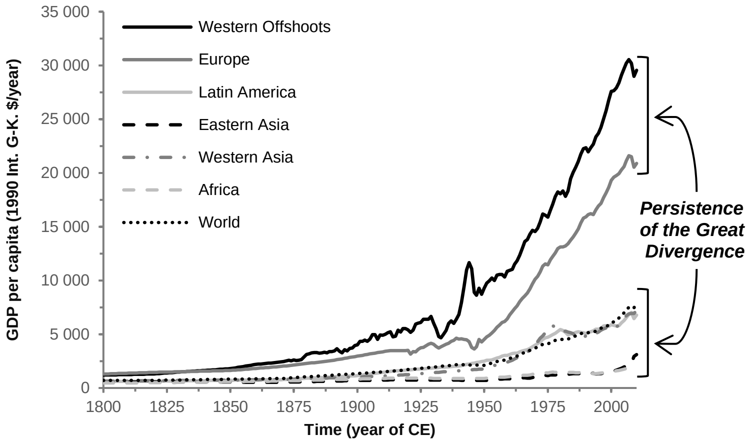


**Figure 1.6 The Great Divergence across regional GDP per capita, 1800–2010 CE. Data source: The Maddison Project (2013).**

The Great Divergence is the second most important fact about economic growth. Understanding its root causes is the second prerogative that any good theory of economic growth should achieve because behind the current (and, of course, not necessary) hegemony of the West lies the most controversial questions of economic growth theory: Is the Great Divergence due to long-term lock-in factors, such as geographic disparities or cultural (or even genetic) predispositions between “Western people and the Rest”, to take Huntington (1996) or Ferguson (2011)’s words? Or, as already mentioned, is it a complex sequence of particular and contingent, and certainly conjectural, events that ultimately gave a lucky head start to Britain in the eighteenth century? If so, why not in the eleventh and twelfth centuries when the economic, industrial and overall technological developments of Eastern China were superior to its European counterparts? Was the Industrial Revolution of the eighteenth century absolutely necessary to escape the Malthusian Trap, or would some other form of development have been possible? Why did regions initially affected by the Great Divergence not converge to narrow inter-country inequalities?

### *1.1.3 Hierarchized-Nested Adaptive Cycles*

#### Civilizational waves/cycles

As already stressed, the largely flat trace of Figure 1.1 is misleading because it does not convey the reality of the successive and often simultaneous phases of development of mankind. In particular, after the rather homogenous hunter-gatherer stage (150,000 to 10,000 BCE), several different civilizations have shaped mankind’s history from the Neolithic Revolution to the present. Defining the word civilization or what the different past and present civilizations are is of course quite perilous. Scholars have argued and will continue to argue about these points for decades. At the risk of considerable oversimplification, we might say that contrary to early states, civilizations encompass large, complex societies with high levels of structural, political, and economic organizations (i.e. the presence of cities, a hierarchy, specialized labor, trade, etc.) accompanied by administrative bureaucracies, and hence writing. Following this definition (enhanced by other items not discussed here for the sake of brevity), Quigley (1961, p.84) defines sixteen past and present civilizations as summarized in Table 1.1. This thesis is not the place to discuss the distinction between the different past and present civilizations,[1] or the reason for their gains and losses of leadership.[2] Yet, it is important to remark that Quigley (1961, p.165) also identifies seven successive phases of civilization development: mixture, gestation, expansion, conflict, core empire, decay, and invasion. With such an analysis comes the idea that the process of human civilization is characterized by successive, and definitely overlapping cycles/waves.

**Table 1.1 Successive civilizational waves, adapted from Quigley (1961).**

| **Civilization name** | **Beginning─end dates** | **Core empire** | **Major invaders** |
|---|---|---|---|
| Mesopotamian | 6000─3000 BCE | Persian | Greeks |
| Egyptian | 5500─300 BCE | Egyptian | Greeks |
| Indic | 3500─1500 BCE | Harappa | Aryans |
| Cretan | 3000─1100 BCE | Minoan | Dorians |
| Sinic | 2000 BCE─400 CE | Han | Huns |
| Hittite | 1900─1000 BCE | Hittite | Phrygians |
| Canaanite | 2200─100 BCE | Punic | Romans |
| Classical | 1100 BCE─ 500 CE | Roman | Germans |
| Mesoamerican | 1000 BCE─1550 CE | Aztec | Europeans |
| Andean | 1500 BCE─1600 CE | Inca | Europeans |
| Hindu | 1500 BCE─1900 CE *(?)* | Mogul | Europeans |
| Islamic | 600─1940 CE *(?)* | Ottoman | Europeans |
| Chinese | 400─1930 CE *(?)* | Manchu | Europeans |
| Japanese | 100 BCE─1950 CE (?) | Tokugawa | Europeans |
| Orthodox | 600 CE─ | Soviet | ? |
| Western | 500 CE─ | *USA*? | ? |

Note: Question marks in italics for Hindu, Islamic, and Chinese civilizations have been added to the original table of Quigley (1961). The *USA* has been added as the probable core empire of the western civilization.

## Secular cycles

Even a partial study of ancient history (such as mine) shows how much, within all major *civilizational cycles*, rises and falls of states and empires have alternated in *secular cycles* of a few centuries. The Hellenistic historian Polybius (200–118 BCE) was probably one of the first to mention such successive secular cycles, but the first clear analysis of such a pattern is generally attributed to the Muslim historian Ibn Khaldun (1332–1406 CE). Far more recently, Turchin & Nefedov (2009) have provided an insightful analysis of the 350 BCE–1922 CE European secular cycles. Based on the demographic-structural theory of Goldstone (1991),

---

[1] Braudel (1995 [1987]) would probably be a good reference to start such an enterprise.

[2] Theories have been formulated to explain the rise of complexity in civilizations, and different authors have tried to explain the apparently unavoidable collapse of successive civilizations by appealing to various (combinations of) factors: depletion of a vital resource, occurrence of an insurmountable catastrophe, invaders, class conflict, elite mismanagement, economic factors, etc. In his masterpiece *The Collapse of Complex Societies*, Tainter (1988) provides an extensive summary of these approaches and proposes the only general satisfactory theory on the rise-and-decline of civilizational complexity, namely that: (i) human societies are problem-solving organizations; (ii) sociopolitical systems require energy for their maintenance; (iii) increased complexity carries with it increased (energy) costs per capita; and (iv) investment in socio-political complexity as a problem-solving response ultimately reaches a point of declining marginal returns above which collapse is just a matter of time. The insightful work of Tainter (probably the best book I have ever read) would really deserve far more development than the few lines I give it here.

these authors argue that a secular cycle can be ideally divided into an *integrative* and a *disintegrative* phase, each being further divided into two parts, *expansion* and *stagflation* on the one hand, *crisis* and *depression* on the other. During the expansion phase, population growth is vigorous, prices and wages are stable, and unified elites support a strong state that can extend its territory at the expense of weaker neighbors. As the population density approaches the carrying capacity of the environment, prices and land rents increase whereas laborers wages decline. Hence, in this stagflation phase, a majority of commoners experience increasing difficulties as growing elites further squeeze the production surplus. Such a situation is untenable and discrete events such as pandemics, extreme episodes of famine, civil wars, and external conflicts become so numerous that they describe a long-lasting state of crisis in which the population declines. In such circumstances, the output per capita increases but intra-elite conflicts continue to generate internal instability. Thus, the crisis grades smoothly into a depression phase that lasts until the ranks of the elites are pruned and population growth between civil war episodes exceeds the declines caused by these same events. Of course, there is no reason to observe a periodicity from one secular cycle to another. The durations of the four phases of each secular cycle are modified by endogenous factors (land-owning and judicial systems, political organization, etc.), exogenous factors (climate, intruders, etc.), and the free will of individuals.

### Generational cycles

Furthermore, Turchin & Nefedov (2009) observed father-and-sons dynamics within the different, and in particular disintegrative, phases of each secular cycle. Indeed, at the beginning of conflicts each act of violence triggers chains of revenge that mean conflicts escalate, but at some point participants lose interest in such atrocities and an increasing number of people yearn for a return to stability. The prevailing social mood swings in favor of cessation of conflict even though the fundamental causes that brought it about in the first place are still operating. The peaceful period lasts for a generation (twenty to thirty years) when the people in charge are the ones who experienced civil war during their youth and are now immunized against it. Eventually, however, the conflict-scarred generation is replaced by a new cohort that did not experience the horrors of civil wars, so that long-lasting unresolved issues can lead once again to internal hostilities. Such alternating social mood dynamics have also been noted by Adams (1891) and Schlesinger (1986) to account for the economic and political history of the USA. According to these authors, US history is a contest between liberalism and conservatism, and its politics can be broadly portrayed as a pendulum swinging between them. Cycles of twenty/thirty years representing the national mood followed one another with a phase of dominant public interest, a transition phase, and then a phase of prevalent private interest. Strauss & Howe (1997) went even further in the theoretical formalization of these *Generational Cycles*. Focusing again on the USA, these authors identify a four-stage social or mood cycle that matches precise generational behavior. These twenty/twenty-five year' turn-arounds are called *High*, *Awakening*, *Unravelling*, and *Crisis*. The description of these four turning points makes it clear that they could be aggregated as successive integrative (*High* and *Awakening*) and disintegrative phases (*Unravelling* and *Crisis*).[1]

### Kondratieff, Kuznets, Juglar, and Kitchin cycles

*Kondratieff cycles*, also called K-waves are probably more famous than the civilizational/secular/generational embedded cycles previously presented. These fifty-year cycles are usually decomposed into an upswing (i.e. integrative) A-phase followed by a downswing (i.e. disintegrative) B-phase. Kondratieff (1935) thought first to appeal to capital investment dynamics to explain the recurrent fifty-year pattern he observed in different indicators (prices, interest rates, heavy industry production quantities, etc.). Scholars have argued a lot about the origin of K-waves, some even claiming that they did not exist. Nowadays, the dynamics of Kondratieff cycles are attributed to major technological waves, i.e. major technological innovations that gradually reshape the world economy but eventually reach a saturation point that can only be overcome by a new wave. Recent empirical evidence shows that K-wave patterns are identifiable in gross world product (GWP) dynamics (Korotayev & Tsirel 2010) and in global patent activity (Korotayev et al. 2011). Those results have converged to the identification of five successive technological Kondratieff cycles since the Industrial Revolution, as summarized in Table 1.2.

**Table 1.2 Successive Kondratieff waves, adapted from Korotayev & Tsirel (2010).**

| Number | Phases | Beginning date | End date | Technological basis |
|---|---|---|---|---|
| I | A: upswing/integrative | 1790 | 1815 | Cotton spinning and weaving. Steam engine (in particular for coal extraction). |
| | B: downswing/disintegrative | 1815 | 1845 | |
| II | A: upswing/integrative | 1845 | 1870 | Steel production and railway transport. |
| | B: downswing/disintegrative | 1870 | 1895 | |
| III | A: upswing/integrative | 1895 | 1915 | Heavy engineering (internal combustion engine, ICE, chemistry) and electricity. |
| | B: downswing/disintegrative | 1915 | 1945 | |
| IV | A: upswing/integrative | 1945 | 1970 | Petrochemical, electronics and automation (mass production). |
| | B: downswing/disintegrative | 1970 | 1990 | |
| V | A: upswing/integrative | 1990 | 2008? | Information and communication technologies (ICTs). |
| | B: downswing/disintegrative | 2008 | ? | |

Note: Dates are arbitrarily precise but there is no doubt that in reality distinction between successive A/B-phases and K-waves are much more blurred. In fact, scholars do not even agree on the exact technological basis of each K-wave and on their causal relations with major economic and geopolitical events.

Suspicions are deeper about the existence and explanations of minor economic cycles that are however worth mentioning since Korotayev & Tsirel (2010) observed them in the spectral analysis they performed on the GWP from 1870 to 2007. *Kuznets cycles* or swings of

---

[1] In the *High* stage, institutions are strong and individualism is weak. Society is confident about where it wants to go collectively (e.g. post-World War II period beginning in 1946 and ending with the assassination of President John F. Kennedy in 1963, Baby Boom Generation). *Awakening* is an era when institutions are attacked in the name of personal and spiritual autonomy. Just when society is reaching its high tide of public progress, people are suddenly tired of social discipline and want to recapture a sense of personal authenticity (e.g. US campus and inner-city revolts of the mid-1960s to the reelection of Ronald Reagan in the mid-1980s; Generation X). During *Unravelling* periods, institutions are weak and distrusted, while individualism is strong and flourishing (e.g. the "US Culture War", beginning in the mid-1980s and ending in the late 2000s; Millennial Generation). *Crisis* is an era in which institutional life is destroyed and rebuilt in response to a perceived threat to the nation's survival. Civic authority revives, cultural expression redirects towards community purpose, and people begin to see themselves as members of a larger group (e.g. current period since the 2000s; Homeland Generation).

15–25 years are connected to major investment in fixed capital for large energy, transport, or communication systems. *Juglar cycles* of about 7–11 years correspond to the replenishment of depreciated capital, e.g. renovation of production machinery. Finally, *Kitchin cycles* appear in the fluctuations of companies' inventories and are generated by time lags in the processing of market information by firms.

Panarchy theory

What the previous paragraphs show is clearly that far from being linear and monotonous, the human adventure is in fact made of patterns that repeat themselves at different space and time scale. It seems pretty obvious that the different cycles/waves that scholars describe independently have in fact a similar structure and, surely, interlinked dynamics. This absence of connection between one type of cycle and another vanishes in light of the *Panarchy Theory* of Gunderson & Holling (2001) synthetized in Holling (2001). This theory was first developed to describe complex natural ecosystems (such as forests) but their authors ingeniously adapted it to social systems. According to Gunderson & Holling (2001), an *adaptive cycle* represents the successive states of any complex system in its three descriptive dimensions: (i) its *potential* for possible future change; (ii) its *connectedness*, describing its level of flexibility and hence internal controllability; (iii) its *resilience*, measuring its vulnerability to unexpected and unpredictable shocks. As shown in Figure 1.7a, the trajectory of a complex adaptive system alternates between long periods of accumulation and transformation of resources (from exploitation phase *r* to conservation phase *K*) with shorter periods that create opportunities for innovation (from release $\Omega$ to reorganization $\alpha$). Cross-scaled hierarchized and nested adaptive cycles form a *panarchy*, as presented in Figure 1.7b and resumed in Table 1.3 for the human adventure.

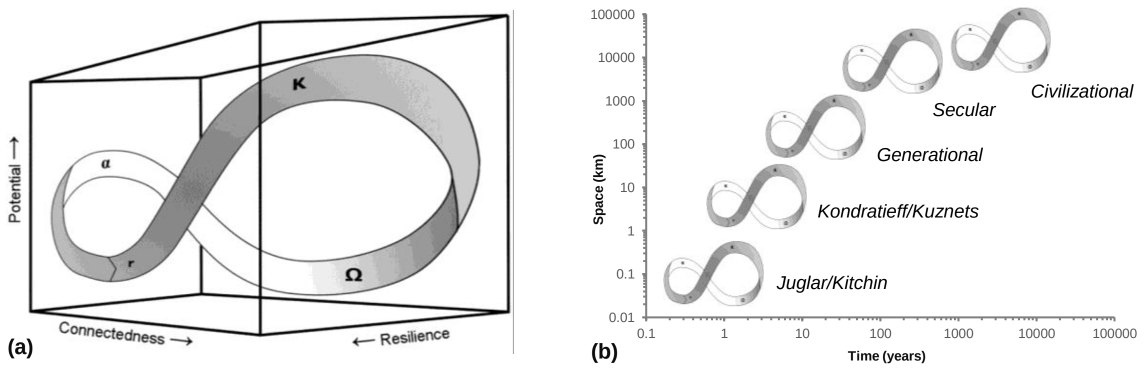


**Figure 1.7 The human adventure as hierarchized-nested adaptive cycles.**
**(a) Typical adaptive cycle in three dimensions. (b) Panarchy of the human adventure. Picture source: Ullah et al. (2015).**

This panarchy represents the fractal structure of the human adventure. In a panarchy, the higher the adaptive cycle, the larger the space scale of the processes involved, and the slower the period of revolution. Moreover, it is important to stress that each level of the panarchy is potentially connected to the next lower and upper stages. When a level in the panarchy enters its Ω phase of creative destruction, the crisis can cascade to the next upper (larger and slower) level. Such an event is most likely if the upper (larger and slower) level is in its $K$ phase, because at this point resilience is low and the level is particularly vulnerable (e.g. an ultimate generational crisis that precipitates an upper secular cycle in crisis phase). In the same way, once a crisis is triggered at one level, the opportunities for, or constraints against, the renewal of the cycle in the $\alpha$ phase are strongly influenced by the $K$ phase of the next slower and larger level (e.g. the saturation of a technological opportunity in a Kondratieff cycle constrains the Kuznets cycle beneath).

In Table 1.3 I have added an upper (and slower) adaptive cycle called the *Energeia cycle*. This larger adaptive cycle is supposed to describe the evolution of the dominant technologically advanced species on Earth, which is currently *Homo sapiens*. The very existence of this adaptive cycle is highly questionable because, as far as we can tell, we are still in the integrative phase of the first of its kind (if such a hypercycle even exists). Consequently, the content of the disintegrative phase of this ultimate upper cycle can only be purely speculative, and we cannot be sure either whether this Energeia cycle will have replicas. The name of this upper adaptive cycle was chosen because, as exposed in Section 1.2 and at more length in Chapter 3, the amount of energy captured and dissipated by any living system provides the first-order description of its level of complexity.

The hierarchized-nested adaptive cycle dynamics of the human adventure is the third most important fact of (very) long-term economic growth. The reasons and consequences of this fundamental structural-functional state are something that any good theory of economic growth has to explain. The rest of this thesis will not deal with this last point that will, however, require some attention in future research. The rest of this thesis will concentrate on explaining the first two facts of long-term economic growth and their link with the long-term pattern of technological change and energy consumption.

**Table 1.3 Hierarchized nested adaptive cycles of the human adventure.**

| Specificity | Name | Periodicity | Basis | Examples | Integrative phase | | Disintegrative phase | |
|---|---|---|---|---|---|---|---|---|
| | | | | | *Exploitation (r)* | *Conservation (K)* | *Release (Ω)* | *Reorganization (α)* |
| Unique? First of many? | Energeia. | Thousands to millions of years. | Life evolution (speciation)/ Energy capture. | Species *Homo sapiens* of the genus *Homo*. | Hunter-gatherer to Agrarian societies. Slow population and technological growth. Low inequalities | Industrial societies. High technological growth. Population progressively stabilizes. Large inequalities. | ? | ? |
| Characteristic of all growth regimes (until proven otherwise). | Civilizational. | Several centuries to (few) thousands of years. | Problem-solving needs and diminishing returns. | Mesopotamian, Egyptian, Sinic, Hittite, Roman Classical, Mesoamerican, Andean, Hindu, Islamic, Western. | Increasing differentiation in structure and functioning. Sprawl of common culture and values. | Increasing bureaucratic and military burdens. Decreasing resilience to inner and outer threats. | Increasing distrust in elites. Loss of cultural integrity. Peripheries disconnecting from core. Collapse. | Simplification or absorption by stronger neighbors. Return to old doctrines and values. |
| | Secular. | 3─8 centuries. | Demographic-structural. | Chinese Imperial Dynasties (Qin, Han, Jin, Sui, Tang, Yuan, Ming, Qing). Western Medieval Europe (Plantagenet, Capetian). | Vigorous population growth. Relatively stable prices and wages. | Increasing difficulties for commoners. Golden age for elites. | Repeating crises: pandemics, famines, civil wars. Population decline. | Elite purge. Larger intercepts between consecutive crises. Population decrease stops. |
| | Generational | 40─45 years<br>(2 generations)<br>80─90 years<br>(4 generations). | Psychology-history contingency. | Intra-secular father-and-sons cycles. Adams/Schlesinger pendulum. Strauss-Howe four turning points. | Strong institutions. Public purpose is dominant, individualism is weak. | Institutions are criticized. Individualism regains interest. Cultural renewal. | Institutions are weak and distrusted. Individualism is flourishing. Society atomized. | Local community is crucial. New spiritual agenda. Institutions are reshaped. |
| *A priori* specific to the Post-Malthusian Regime and Modern Growth Era. | Kondratieff. | 40─60 years. | Technological. | Cotton and steam engine, steel and railway, electricity and chemistry, petrochemical and automation, ICT. | Technological breakthrough, flurry of applications and improvements. | Widespread adoption and diffusion. Optimization. | Saturation of application and possibilities of improvements. | Research and Development (R&D) clustering. Look for new innovative breakthrough. |
| | Kuznets. | 15─25 years. | Infrastructural investment. | Installation of large energy, transport, and communication system. | Increasing infrastructure deployment. | Maturity/Saturation of infrastructure deployment. | No additional infrastructure deployment. | Optimization of vintage infrastructure. |
| | Juglar. | 7─11 years. | Capital replenishment. | Compensation of machine depreciation, or more speculatively, real estate bubble. | Fast capital replenishment. Capital is always brand new. | Slower capital replenishment. Vintage capital appears. | Low capital replenishment. Vintage capital is substantial. | No capital replenishment. Capital depreciate substantially. |
| | Kitchin. | 3─5 years. | Information time lags. | Inventory adjustment to market mood, critical metal shortage. | Stocks accumulate rapidly. | Stocks stabilize slowly at maximum. | Stocks decrease rapidly. | Stocks stabilize slowly at minimum. |

## 1.2 TECHNOLOGICAL CHANGE AND ENERGY CONSUMPTION

### *1.2.1 Technological Change and Total Productivity Factor*

Technological change is essential to understanding long-term economic growth but correctly defining this term is perhaps one of the most difficult assignments that an economist can be given. In standard growth models, new technologies cause growth by increasing the amount of output that can be produced from a given set of resources (Lipsey et al. 2005, p.10). Basically, it means that if one observes that people can produce more output (such as GDP) with fewer inputs (such as physical and human capital, and routine labor), it means that technology has changed to improve the efficiency of the production apparatus. As developed in Appendix A, technological change is called *Total Factor Productivity* (TFP) or *Solow Residual* under such an approach as it is the remainder of the difference between observed increases of output and inputs. Such a view has three related, and paradoxical, problems. (i) Aggregated technological change appears as a continuous process that is, however, divided into three stages at the micro level, namely: invention, innovation, and diffusion. (ii) No distinction is made between, on the one hand *Usherian* technological change made of small incremental changes that tend to optimize existing technologies, and *Schumpeterian* technological change on the other hand that can possibly reshape the whole economy thanks to discontinuous and radical inventions which render obsolete existing and older technologies, capital, and possibly labor skills.[1] (iii) Very different production-augmenting factors are grossly aggregated in this single variable labeled TFP, including: the primary-to-final and final-to-useful energy conversion efficiency (if energy is considered as an input factor), the division and organization of labor, the broader organization and efficiency of markets, the skill improvements of laborers, the contribution of information and communication technologies, but also the beneficial effects of inclusive institutions (which, for example, protect private property rights and consequently incentivize innovation and R&D). I will comment at more length on these facts later on but for the time being let us note that different references are worth studying to grasp the many aspects of technological change; those include Mokyr (1990), Grossman & Helpman (1991a), Helpman (1998), and Lipsey et al. (2005). In all these references the clear link between major technological breakthroughs and the historical energy consumption pattern of mankind is largely ignored.

### *1.2.2 General Purpose Technology and Energy Dependence*

Major technological breakthroughs that have a particular influence on the economy have received different names (e.g. Mokyr’s “macro inventions”) but the term *General Purpose Technology* (GPT) now seems settled. According to Lipsey et al. (2005, p.98), a GPT “is a single generic technology, recognizable as such over its whole lifetime, that initially has much scope for improvement and eventually comes to be widely used, to have many uses, and to have many spillover effects”. Those authors further stress that GPTs are typically use-radical but not

---

[1] I took the Usherian/Schumpeterian distinction from Ayres & Warr (2009, p.17).

technology-radical, meaning that GPTs do not stand out from other technologies because of a revolutionary technological basis, but rather because of outstanding applications and adaptations to other technologies and sectors of the economy. GPTs are typically not born in their final form, so they often start off as something we would never call a GPT and then develop into something that transforms an entire economy (Lipsey et al. 2005, p.97). The considerable scope of improvement of GPTs is explored as their range and variety of use increase, which in the meantime generate knowledge and practical spillovers on other technologies and organizational processes. Table 1.4 gives a list of historical transforming GPTs adapted from Lipsey et al. (2005, p.132).

**Table 1.4 Transforming GPTs, adapted from Lipsey et al. (2005).**

| No. | GPT's name | Date of widespread use | Classification |
|---|---|---|---|
| 1 | Stone, bone, and wood tools [a] | Before 150,000 BCE [b] | Material |
| 2 | Mastery of fire [a] | Before 150,000 BCE | Energy |
| 3 | Domestication of plants | 9000─8000 BCE | Energy |
| 4 | Domestication of animals | 8500─7500 BCE | Energy |
| 5 | Smelting of copper ore | 8000─7000 BCE | Material |
| 6 | Wheel | 4000─3000 BCE | Transport |
| 7 | Writing | 3400─3200 BCE | Information |
| 8 | Bronze | 2800 BCE | Material |
| 9 | Iron | 1200 BCE | Material |
| 10 | Waterwheel | Early medieval period | Energy |
| 11 | Three-mastered sailing ship | 15th century | Transport |
| 12 | Printing | 16th century | Information |
| 13 | Steam engine | Late 18th to early 19th century | Energy |
| 14 | Factory system | Late 18th to early 19th century | Organization |
| 15 | Railway | Mid 19th century | Transport |
| 16 | Iron steamship | Mid 19th century | Transport |
| 17 | Internal combustion engine | Late 19th century | Energy |
| 18 | Electricity | Late 19th century | Energy |
| 19 | Petrochemistry [a] | 20th century | Material |
| 20 | Mass production (with continuous process) | 20th century | Organization |
| 21 | Computer | 20th century | Information |
| 22 | Internet | 20th century | Information |
| 23 | Biotechnology/Nanotechnology [c] | Sometime in the 21st century | Material/Information |

[a] Quite astonishingly, Stone Age tools, mastery of fire and petrochemistry are not mentioned in the original survey of Lipsey et al. (2005).

[b] *Before 150,000 BCE* is used because defining a date of widespread use appears impossible for Stone Age tools and mastery of fire. Those technologies (not listed in Lipsey et al. 2005) were even used by hominid, such as *Homo habilis* and *Homo erectus*, which preceded *Homo sapiens*.

[c] Of course, the GPT statuses of biotechnology and nanotechnology are for now purely speculative and remain to be confirmed.

The most striking fact of Table 1.4 is that, even if only seven out of the total 23 GPTs of this list are *directly* energy-related, all 16 other GPTs necessitate or imply *indirectly* increasing energy consumption to deserve their GPT status. Mastery of fire, the domestication of plants and animals, waterwheels, steam engines, internal combustion engines, and electricity are different technologies that *have directly implied an increase in the level of energy consumed* by societies. Transport GPTs (the wheel, sailing ships, railways, and iron steamships) *are* *naturally used in combination with energy* to propel people, goods, and information. Material GPTs (rudimentary tools, iron and bronze smelting and working, and petrochemistry) *inevitably require energy*. The huge reshaping of the economic process brought by organizational GPTs (factory system, mass production) *can only be achieved with equally substantial energy consumption*. To an equal extent, the spread of informational GPTs (writing, printing, computing, and internet) *is necessarily supported by increasing energy capture*.

### *1.2.3 Energy Consumption and Societal Development*

#### Foragers, farmers, and industrial man

The fact that the level of development of societies is closely linked to their level of energy consumption has never been so well highlighted as in the words of White (1943) and in the graph of Cook (1971) reproduced in Figure 1.8 with amendments (initial values were in kcal/day/capita and associated major energy-related breakthroughs were not present in the original).

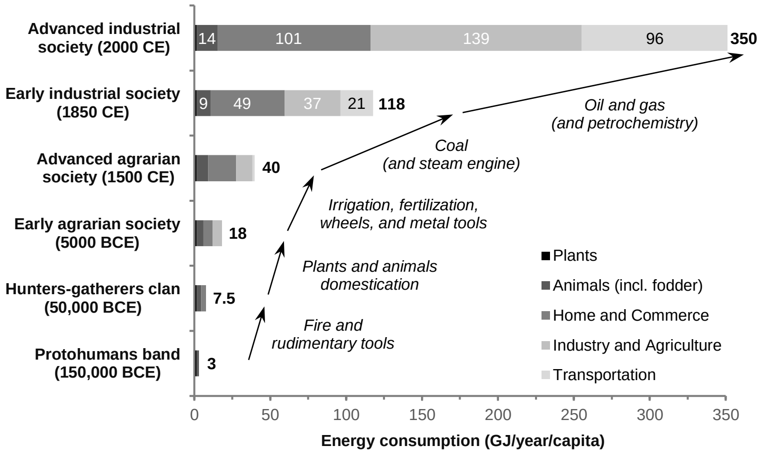


**Figure 1.8 Energy consumption at different stages of societal development.**
**Data source: Cook (1971), originally in kcal/day/capita without associated major energy-related breakthroughs.**

For 92.1% of its history (150,000–10,000 BCE) mankind has been organized into highly egalitarian hunter-gatherer clans within which each individual only consumed 7.5 GJ/year, mostly in the form of food and to a lesser extent woodfuel. Stone tools and the use of fire could not allow a greater energy capture, so that material property during this period was rather rudimentary compared to modern standards. The domestication of plants and animals induced an almost three-fold increase in the average energy consumption per capita (18 GJ/year), which allowed the establishment of permanent settlements and the premises of the division of labor division and political hierarchization.[1] Power delivered by draft animals (and fed with fodder, i.e. sunlight energy converted into phytomass) contributed substantially to this pattern. The gradual improvement of metal tools, the increasing use of organic fertilizers (manure), and the introduction of new farming technologies (irrigation, the wheel) induced a further two-fold

---

[1] This sentence could misleadingly suggest that the domestication of plants and animals preceded and so caused sedentary lifestyles. The evolution from foraging to farming was in fact far more gradual, and (almost) permanent villages of hunter-gatherer were established in the Hilly Flanks before plants and animal were domesticated (Diamond 1997; Morris 2010).

increase in the average energy capture per capita (40 GJ/year). Societal development at that stage was represented by bigger and more connected city-states with increasing military power which could even sustain large empires. In total, the time for which the farming-energy-capture system was representative of most of the global population represented 7.8% of mankind's history (10,000 BCE–1850 CE). The last 0.1% (1850 CE–2016 CE), which has embedded the most colossal changes in societal development, has been associated (and in fact largely caused/allowed) by the opportunity of mankind to tap into fossilized solar energy in the form of coal, oil, and gas. The average western individual now controls 350 GJ/year, a seven-fold increase compared to the pre-1800 protoindustrial individual, enabling individuals to transform and transport many more materials and convey much more information.

Energy transitions in the last two centuries

Figure 1.9a,b presents the evolution of the use of the *Grand Chain of Energy* (as coined by Morris 2010), at global scale by *industrial man* between 1800 and 2010 CE. The data and methodology used to build these graphs are detailed in Appendix B. Rather clearly, the pre-industrial global energy consumption mix (15% food, 10% fodder, 1% traditional water/wind, and 74% biomass) has been altered dramatically since the Industrial Revolution. In a first step, coal largely replaced woodfuel and allowed the widespread use of steam engines for railways, steamships, and industry. On the eve of World War I, coal reached its maximum share of 49% of the global primary energy consumption mix. In a second step, crude oil found even more applications and allowed tremendous efficiency gains and cost reductions. One of the most famous examples of this fact is the decision made by Winston Churchill to convert the entire British fleet from coal to oil in 1914, which gave Britain the fastest navy in the world and a consequent decisive advantage over Germany in this time of war. After the two oil crises of the 1970s, various countries tried to reduce their dependence on crude oil, so that gas production, nuclear electricity, and hydropower increased considerably. So-called clean technologies which many experts see as the future of mankind, such as wind turbines, solar panels, tidal and wave electricity, currently represent 1% of the global primary energy supply (and 10% of its renewable part, woodfuel and crop residues still contribute for 70%, and hydro the remaining 20%).

The more than close relation between major technological changes (such as GPTs) and the ever-increasing energy consumption pattern that has accompanied the human adventure is the fourth main fact of long-term economic growth. Understanding the exact role of energy in the economic process must surely be at the root of any good theory of economic growth. Unfortunately, for reasons that will be set out at the beginning of Chapter 3, energy has been completely discarded from standard economic growth theories. A logical preliminary step to understanding this is to review all existing theories of economic growth in order to see how scholars explain the growth process without referring to energy. In the following sections I present the narrative-based models proposed by historians (and supported by econometricians). Mathematics-based models advanced by economists who concentrate on the mechanisms of the Modern Growth Era will be investigated in Chapter 2.

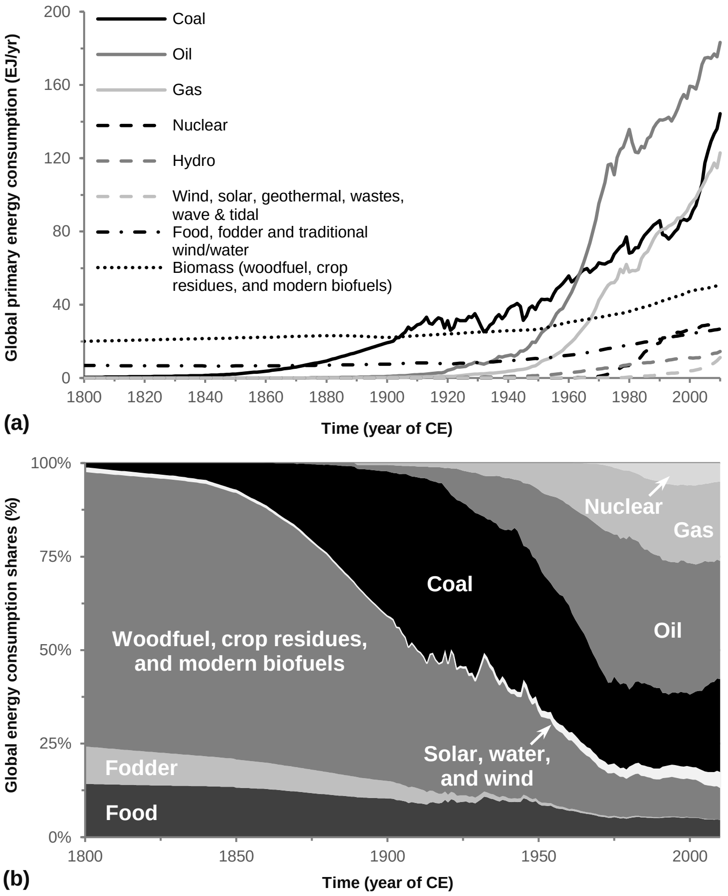


**Figure 1.9 Global primary energy consumption, 1800–2010 CE.**
**(a) Annual quantities (EJ/year). (b) Relative shares (%). "Solar, water and wind" aggregate traditional waterwheels/windmills with renewable electricity producing technologies (i.e. wind power, solar PV, solar thermal, geothermal, wastes, wave, tidal, and OTEC). Data source: see Appendix B.**

## 1.3 DEEP-ROOTED CAUSES OF ECONOMIC GROWTH

Historians generally seek to explain the current Western hegemony (in terms of wealth, technology, and military power) through deep-rooted factors and mechanisms within the Malthusian Epoch (Eurocentric group) or through short-term accidental events of the Post-Malthusian Regime (California school). Long-term lock-in theorists of the Eurocentric group appeal to three kinds of deep-rooted hypothesis, namely biogeographical endowment, cultural factors, and institutional determinants. Short-term accidental theorists of the California school highlight that regions of the world had many more similarities than differences on the eve of the Industrial Revolution, and that the reasons for the British appropriation of this turning event was essentially a matter of conjuncture and luck.

### *1.3.1 Biogeographical Hypothesis*

Local climate, sea access, and diseases

Theories belonging to the *Biogeographical Hypothesis* suppose that favorable biogeographical conditions fostered the earlier Western European take-off from the Malthusian Epoch and explain the divergence in income per capita around the globe. The earliest[1] version of such thinking is generally attributed to the socio-political climate theory of Montesquieu (1748), according to which:

> You will find in the climates of the north, peoples with few vices, many virtues, sincerity and truthfulness. Approach the south, you will think you are leaving morality itself, the passions become more vivacious and multiply crimes. . . . The heat can be so excessive that the body is totally without force. The resignation passes to the spirit and leads people to be without curiosity, nor the desire for noble enterprise.

The basic idea here, also found in Marshall (1890, p.195) and Toynbee (1934), is that hot and wet climates are detrimental to hard work and creativity, and furthermore imply little effort in providing shelter and gathering food, whereas cold and dry climates are conducive to and necessarily require, much more work and ingenuity.

Quite differently, Braudel (1996 [1949]) emphasizes the key role of Mediterranean and North Atlantic coastal countries as the creative centers of global capitalism after the fifteenth century. McNeill (1963) similarly stresses Europe’s great advantages in coastal trade, navigable rivers, temperate climate, and disease patterns as fundamental conditions for its take-off and eventual domination of the Americas and Australia. Jones (1981) and Crosby (1986) have placed geography and climate at the center of their explanations for Europe’s preeminent success in economic development. Far from thinking that climate has a mechanical one-to-one relation with economic development, Kamarck (1976, p.11) stresses that, in today’s poor countries, climatic factors have hampered economic development through their impact on agriculture (directly or through the diseases and pests afflicting animals and plants), mineral discovery, and man himself through disease. Bloom & Sachs (1998) detail these points and argue that, in Africa in particular, tropical agriculture is faced with chronic problems of low yields and fragility due to low photosynthetic potential, high evapotranspiration, low and variable rainfall, highly weathered soils, veterinary diseases, and plant and animal pests. For these same authors, evidence suggests that the burden of infectious disease (particularly malaria) is vastly higher in the tropics than in the temperate zones. Furthermore, Bloom & Sachs (1998) support econometrically that the failure of Africa to control diseases is not mainly the result of poor public health measures, unresponsive governments, or poverty, but it is due rather to the natural environment. Finally, these authors also point to Africa’s remarkable disadvantages in transport costs to explain its long-term development lag. Those include: (i) a great distance from major world markets in the northern midlatitudes, in particular, separation from Europe by the vast Sahara desert; (ii) a very short coastline relative to the land area; (iii)

---

[1] The idea of the effect of climate on income through its influence on work effort is in fact traced back to Machiavelli (1997 [1519]) but I find Montesquieu’s wording particularly eloquent.

very few natural coastal ports; (iv) the highest proportion of landlocked states, and the highest proportion of the population within landlocked states, of any continent; and (v) the absence of rivers leading into the interior of the continent that are navigable by ocean-going vessels, as are the Rhine, the Mississippi, the Amazon, and the Yangtze on other continents. The statistically significant impact of geographical endowment (through climate and land openness) on per capita GDP growth is even more consistently demonstrated by Gallup et al. (1999) who conclude that Sub-Saharan Africa is especially hindered by its tropical location, high prevalence of malaria, small proportion of people living near the coast, and low population density near the coast. Europe, North America, and East Asia, by contrast, have been favored on all four counts according to these authors.

### The timing of the agricultural revolution

A radically different version of the *Biogeographical Hypothesis* was presented by Diamond (1997) in his influential and Pulitzer Prize-winning essay *Guns, Germs, and Steel*. Through a backward induction approach, Diamond argues that a few *ultimate* (more properly *deep-rooted* to my mind) biogeographical factors have profoundly predetermined mankind’s history. According to him, if Western Europe rules (for now), it is thanks to technological (guns, large sail ships, higher disease resistance, etc.) and institutional (large markets, cities, writing, political organization, etc.) advantages that were profoundly established circa 1500 CE and explain that Westerners colonized the New World (and not the other way around) and two hundred and fifty years later launched the Industrial Revolution. For Diamond if Westerners had such large technological and institutional advantages circa 1500 CE, it was because Western Eurasia was the first region to experience the Agricultural Revolution of the Neolithic with several millennia of advance compared to other continents, and hence benefited from the early establishment of cities, high population densities, and associated non-food producing elites that created and organized knowledge. As the author argues, if agriculture first emerged in the Hilly Flanks of Southwest Asia and then easily spread to Western Europe and the rest of Eurasia, it is not because their inhabitants were cleverer and better adapted to their environment, it is just because their environment offered them a higher number of suitable plants and animals for domestication as shown in Table 1.5. For example, of the 56 wild large-seeded grass species of the world, 32 were present around the Mediterranean Sea, whereas East Asia only had 6, Mesoamerica 5, Sub-Saharan Africa 4, and South America and Oceania 2. Similarly, out of the world’s 14 domesticated herbivorous mammals weighing more than 45 kg (and hence adapted to agricultural work), 13 were in Eurasia, only 1 in South America, and 0 in Africa and Oceania. With such an uneven distribution of wild plants and animals suitable for domestication, the differential timing of the agricultural onset in different regions of the world was predetermined and could hardly have been altered. The reasons for the unequal distributions of domesticable plants and animals across regions are numerous. First, Eurasia is the largest terrestrial continent, so that other things being equal its biodiversity should be higher than other continents such as Africa, America and Oceania. Second, regarding plants, the temperate climate around the Mediterranean Sea has surely been influential in favoring large-seeded grass species compared to the equatorial and tropical climates of Sub-Sharan Africa, Mesoamerica, and South

America.[1] Third, concerning large mammals, Martin (1967; 1984) posits that the later *Homo sapiens* reached various regions, the greater was their skill as big game hunters and the less experience their prey had with human predators, which resulted in the rapid extinctions of large animals in the Americas and Australia in the late Pleistocene.[2]

**Table 1.5 The different onsets of agriculture, adapted from Olsson & Hibbs (2005).**

| Area | Earliest date of domestication | Domesticated plants (number of large-seeded grass species[a]) | Domesticated animals (number of suitable animals for domestication[b]) |
|---|---|---|---|
| *Independent origin (initial cradle)* | | | |
| Southwest Asia (Hilly Flanks) | 8500 BCE | Wheat, pea, olive. (32) | Sheep, goat (9) |
| East Asia (Yangtze delta) | 7500 BCE | Rice, millet. (6) | Pig, silkworm (7) |
| Mesoamerica (Mexico) | 3500 BCE | Corn, bean, squash. (5) | Turkey (0) |
| Andes and Amazonia (Peru) | 3500 BCE | Potato, manioc. (2) | Llama, guinea pig (1) |
| Eastern United States | 2500 BCE | Sunflower, goose foot. (4) | - |
| (?) Sahel | 4000 BCE | Sorghum, African rice. (4) | Guinea fowl (0) |
| (?) Tropical West Africa | 3000 BCE | African yams, oil palm (4) | - |
| (?) Ethiopia | ? | Coffee, teff (4) | - |
| (?) New Guinea | 7000 BCE? | Sugar cane, banana (0) | - |
| *Following arrival of founder crops from elsewhere* | | | |
| Western Europe | 6000–3500 BCE | Poppy, oat. (dozen) | - |
| Indus Valley | 7000 BCE | Sesame, eggplant. (several) | Humped cattle (1) |
| Egypt | 6000 BCE | Sycamore fig, chufa. (several) | Donkey, cat (1) |

[a]The numbers refer to the geographical distribution of the world's 56 heaviest wild grasses (Blumler 1992). The figures do not add to 56 because some species are found in more than one location.
[b]The numbers refer to the geographical distribution of the world's 14 domesticable herbivorous or omnivorous, terrestrial mammals weighing more than 45 kg (Nowak 1991). The figures do not add to 14 because some species are found in more than one location.

Despite the higher endowment of suitable plants and animals for domestication, Diamond (1997) stresses that two other particular features of Eurasia explain its development advantage, namely (i) the East–West orientation of Eurasia vs. the North–South orientation of America and Africa, and (ii) the higher number of natural barriers (desert, dense forests, and terrestrial bottlenecks such as the Isthmus of Panama) in these last two continents compared to Eurasia. Regarding the first point already advanced by McNeil (1963), the East–West orientation of Eurasia implies far less latitudinal variation compared to other continents, and hence more similar day length, seasonal variations, regimes of temperature and rainfall, and diseases. All these points are conducive to the spread of species, best-practices, and more generally technology across localities. Concerning the second point, Diamond (1997) argues that technologies in the critical areas of agriculture and health could easily diffuse within similar ecological zones, but not across different ecological zones, which explains that economic development spread through the temperate zones but not through the tropical regions. Hence, these two other deep-rooted biogeographical factors have implied an easier widespread

---

[1] Olsson & Hibbs (2005) have shown that exogenous geographic conditions (climate, latitude, continental axis and size) explain around 80% of the variance of the international distribution of heavy seeded plants and large domesticable animals that are known to have existed in prehistory.

[2] This "blitzkrieg model of overkill'' as it is called by Martin (1967; 1984) is fully supported by some simulation models (Alroy 2001), and nuanced by others (Brook & Bowman 2002). See also Grayson (1991) for the alternative climate change-related hypothesis of the Pleistocene megafauna extinction in America and Oceania.

diffusion of agricultural and technological best practices, but also of alphabets and languages, in Eurasia compared to Africa and America.

The earlier onset of agriculture in Southwestern Asia and its rapid diffusion to Europe was a matter of higher probability. Domesticated plants and animals gave first to Western Eurasia a reliable source of food with high nutritional value, but also fertilization, wool, leather, transport, plowing, and military power that could feed a much greater population per unit area and sustain an increasing proportion of non-food-producing but technology-inventing population. Moreover, the close physical proximity of man and animal also gave Eurasian agriculturists a high resistance to animal-related germs such as those causing smallpox, measles, and tuberculosis, which proved to be decisive during the colonization of the New World since germs brought from Europe killed more native Americans than guns and swords. As shown in Figure 1.10 where the technological and organizational trajectories of the different regions of the world are represented, the head start of Western Eurasia lasted for millennia and was slow to resorb.

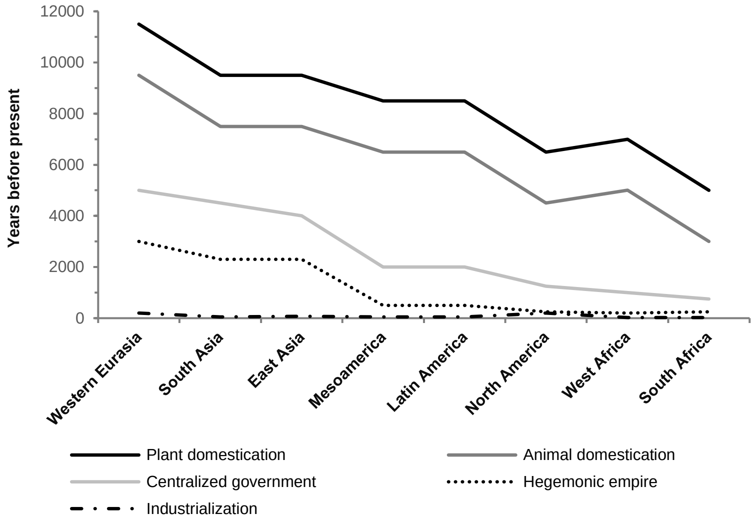


**Figure 1.10 Development trajectories of different world regions, 10,000 BCE–2000 CE. The more horizontal the line, the smaller the technological and organizational gaps between regions, the more uniform the world. Source: reproduced from Morris (2015, p.153).**

The inspiring thesis of Diamond is supported by two econometric studies (that furthermore both propose a theoretical mathematical model of the Malthusian Epoch). Olsson & Hibbs (2005) show that the unequal distribution of domesticable plants and animals accounts for around two-thirds of the regional variation in the estimated dates of the agriculture onset. These authors further show that exogenous geography (climate, latitude, continental axis and size) and initial biogeographical conditions (number of domesticable plants and animals) are able to account for half of the sixty-fold difference in contemporary per capita income observed in a broad international cross-section of 112 countries. Moreover, the geographic and biogeographic signals that these authors have detected in current levels of income per person were robust to controls for political and institutional variables. These results mean that current variations in economic prosperity still significantly embody the effects of prehistoric biogeographical conditions. More recently, Ashraf & Galor (2011) found a highly statistically significant positive effect of regional differences in land productivity and the number of years elapsed since the Neolithic Revolution on regional population density in the years 1 CE, 1000 CE, and 1500 CE. However, according to the same authors, the effects of land productivity and the number of years elapsed since the Neolithic Revolution on the per capita income of the same periods are not significantly different from zero (which contradicts the results of Olsson & Hibbs 2005 on this point). Importantly, the qualitative results remain robust to controls for the confounding effects of a large number of geographical factors, including absolute latitude, access to waterways, distance to the technological frontier, and the share of land in tropical versus temperate climatic zones, which may have had an impact on aggregate productivity either directly, by affecting the productivity of land, or indirectly via the prevalence of trade and the diffusion of technologies. Hence, the analysis of Ashraf & Galor (2011) provides the first test of Diamond's (1997) influential hypothesis in the context of preindustrial Malthusian societies, establishing that, indeed, an earlier onset of the Neolithic Revolution contributed to the level of technological sophistication and thus population density in the pre-modern world.

As pointed out by Acemoglu & Robinson (2012, p.52), although Diamond's argument, and to a lesser extent other versions of the *Biogeographical Hypothesis*, are powerful explanations for intercontinental differential growth, they can hardly elucidate the modern level of economic inequality between countries of the same world region. In other words, although the endowment of biogeographical factors has surely had a significant effect on economic development in the long run, the interplay of other factors is needed to fully explain the economic growth process.

### *1.3.2 Cultural Hypothesis*

The idea that the Western take-off two hundred years ago is due to the unique growth-compatibility of its Christian culture based on reason, inventiveness, and freedom (finding its origin in ancient Greece two thousand years ago), compared to the obscurantism and conservatism of the Confucian East and Islamic Middle East is probably the most popular explanation of the Great Divergence among laypersons. Fortunately, it is hard to find such simplistic views in scholars' works because theories belonging to the *Cultural Hypothesis* try more convincingly to relate cross-country differences in economic growth to cultural variations among populations.

#### Protestant work ethics, entrepreneurship, and the rise of modern science

Those who emphasize the role of culture usually see the long-term as being far shorter than proponents of the *Biogeographical Hypothesis*. Among them, Weber (1930 [1905]) became influential when he stressed that the Protestant Reformation and the *Protestant work ethic* it spurred in the sixteenth century played a key role in the rise of modern industrial society in Western Europe. Weber argues that contrary to Catholicism, Protestantism defined and sanctioned an ethic of everyday behavior that is conducive to business success because the Protestant work ethic makes people work harder, more effectively, and is akin to entrepreneurship. In spite of diverse qualitative rebuttals (Tawney 1926; Samuelsson 1961), econometric studies seem unable to concretely support the Weberian Protestant work ethic theory (Arruñada 2010; van Hoorn & Maseland 2013; Nunziata & Rocco 2016). But another point Weber made was that the Protestant Reformation narrowed the gender gap in school enrollment and literacy rates. Not surprisingly, this social aspect of Weber's theory has found much more support among neoclassical economists who put a lot of emphasis on human capital to explain the economic growth process (Becker & Woessmann 2009; 2010; Schaltegger & Torgler 2010).

Following the same line of thought, some scholars have closely tied the Protestant Reformation to the rise of modern science. De Candolle (1885) counted that of the ninety-two foreign members elected to the French *Académie des Sciences* in the period 1666–1866, some seventy-one were Protestant, sixteen Catholic, and the remaining five Jewish or of indeterminate religious affiliation, this from a population pool outside of France of 107 million Catholics and 68 million Protestants. A similar count of foreign Fellows of the Royal Society in London in 1829 and 1869 showed similar relative proportions of Catholics and Protestants out of a pool in which Catholics outnumbered Protestants by more than three to one. Merton (1938) focused on English Puritanism and German Pietism as being responsible for the development of the scientific revolution of the seventeenth and eighteenth centuries. He explains that the connection between religious affiliation and interest in science is a result of a significant synergy between the ascetic Protestant values and those of modern science. According to this same author, Protestant values encouraged scientific research by allowing science to identify God's influence in the world and thus providing religious justification for scientific research. Focusing originally on the city of Detroit in the late 1950s, Lenski (1961) argued that the contributions of Protestantism to material progress have been largely unintended by-products of certain distinctive Protestant traits. According to this author, the Reformation encouraged intellectual autonomy among Protestants and the differences between Protestants and Catholics have survived to the present day, so that most of today's leading industrial nations are Protestant (USA, UK, Germany). Other scholars such as Jacob (1988; 1997), Mokyr (2005; 2009), and to a lesser extent Goldstone (2009) and Vries (2015) attribute much of the credit for the burst of innovations and accelerated diffusion of best practices after 1750 to the scientific culture of Western Europe, and in particular Britain. They argue that Western European societies were particularly dynamic and inclined to see technological breakthrough in the eighteenth century thanks to the increase or propagation two hundred years before of: high literacy rates, printing, publishers, scientific societies, university networks, relatively accessible public lectures, and so on. For these authors, changes in the intellectual and social environment and the institutional background in which knowledge was generated and disseminated from the sixteenth to the eighteenth centuries explain the success of the Industrial Revolution.

With his book *The Wealth and Poverty of Nations*, Landes (1998) is usually said to be the main proponent of the *Cultural Hypothesis* among contemporary scholars.[1] Speaking about the Inca people, Landes asserts that "in any event, the culture deprived the ordinary person of initiative, autonomy, and personality" (*Ibid.*, p.112). About the early British industrial take-off,

---

[1] With the more recent *Civilization: The West and the Rest*, Ferguson (2011) is probably the most serious contender of Landes to hold such a title.

he stresses to "consider not only material advantages (other societies were also favorably endowed for industry but took ages to follow the British initiative), but also the nonmaterial values (culture) and institutions" (*Ibid*., p.215[1]). Further in his book, Landes argues that if "the Balkans remain poor today", "it was not resources or money that made the difference; nor mistreatment by outsiders. It was what lay inside—culture, values, initiative. These peoples came to have freedom enough. They just didn't know what to do with it" (*Ibid*., pp.252–253). Regarding Islamic culture, Landes thinks it: "(1) does not generate an informed and capable workforce; (2) continues to mistrust or reject new techniques and ideas that come from the enemy West (Christendom); and (3) does not respect such knowledge as members do manage to achieve, whether by study abroad or by good fortune at home. At the most elementary level, the rates of illiteracy are scandalously high, and much higher for women than for men. That alone speaks of a society that accords women an inferior place, and this is clearly related to attitudes cultivated in Islam and especially in the Islam of the Arab world" (*Ibid*., p.410). His judgment is that "if we learn anything from the history of economic development, it is that culture makes all the difference. (Here Max Weber was right on.)" (*Ibid*., p.516). Nevertheless, it would be unfair to say that Landes thinks that culture alone explains all the differences among countries' abilities to generate wealth. "Economic analysis cherishes the illusion that one good reason should be enough, but the determinants of complex processes are invariably plural and interrelated" (*Ibid*., p.517).

### Trust, religion, and religiosity

So far, I have mostly presented ideas from historians and political scientists without clearly defining the notion of culture, but such a definition is required when it comes to seeing how econometric studies can support the *Cultural Hypothesis*. Guiso et al. (2006, p.23) define culture as "those customary beliefs and values that ethnic, religious, and social groups transmit fairly unchanged from generation to generation". Because such a definition of culture is hardly quantifiable, culture entered the economic (and econometric) discourse through the concept of trust, defined by Gambetta (1988) as "the subjective probability with which an agent assesses that another agent or group of agents will perform a particular action". Several econometric studies (Knack & Keefer 1997; Zak & Knack 2001; Dincer & Uslaner 2010) demonstrate that trust and civic cooperation have a significant positive correlation with aggregate economic activity, an idea developed by different political scientists such as Banfield (1958), Putnam (1993), and Fukuyama (2005). Regarding religion, Barro & McCleary (2003) show that for given religious beliefs, increases in church attendance tend to reduce economic growth. In contrast, for given levels of church attendance, increases in some religious beliefs, notably belief in hell, heaven, and an afterlife, tend to increase economic growth. Barro & McCleary's (2003) conjecture is that stronger religious beliefs stimulate growth because they help sustain specific individual behaviors (such as honesty, a work ethic, thrift, trust, and openness to strangers) that enhance aggregated economic productivity. And indeed, Guiso et al. (2003) show that being raised religiously raises the level of trust by 2 percent, whereas regularly attending religious services increases trust by another 20 percent compared to nonreligious

---

[1] On this same page, Landes gently mocks the reluctance of scholars to use the words *culture* or *values*, as shown by Rostow's (1963) "propensities" and Abramowitz's (1989) "social capability".

people. Furthermore, Guiso et al. (2003) find that on average Christian religions are more positively associated with attitudes that are conducive to economic growth (trust in others and the legal system, respect of women rights), while Islam is negatively associated. The ranking between the two main Christian denominations is less clear. It appears that Protestants trust others and the legal system more than Catholics, and they are less willing to cheat on taxes and accept a bribe. By contrast, Catholics support private ownership twice as much as Protestants and are more in favor of competition than any other religious group (including Protestants). Ethnic origin also seems to have an impact since Guiso et al. (2006) find that the level of trust a US citizen has toward others depends in part upon where their ancestors came from.

### Ethnic, linguistic and religious fractionalization

Another important body of studies of the *Cultural Hypothesis* concerns the impact of the level of ethnic, linguistic and religious fractionalization on economic growth. Knack & Keefer (1997) have supposedly shown that trust and norms of civic cooperation (that positively affect economic growth) are stronger in countries that are less polarized along lines of class or ethnicity. Similarly, Easterly & Levine (1997) assert that cross-country differences in ethnic diversity are positively correlated to a substantial part of the cross-country differences in public policies, political instability, and other economic factors associated with long-run growth. Arcand et al. (2000) harshly criticize the methodology employed by Easterly & Levine (1997), while Alesina et al. (2003) nuance their results. The latter explain that ethnic and linguistic fractionalization variables are likely to be important determinants of economic success, but that strong correlation with other potential variables, in particular geographical ones, greatly complicates the evaluation of the size of these effects. Collier (2000) and Alesina & La Ferrara (2005) argue that fractionalization has negative effects on growth and productivity only in nondemocratic regimes, while democracies manage to cope better with ethnic diversity.

As can been seen in this literature review, the problem with the *Cultural Hypothesis* lies in the difficulty in establishing a straightforward causal link between core belief and preferences on the one hand, and economic performances on the other. All the econometric results previously cited are based on multiple linear regressions (that most of the time use proxies to control for geographical and institutional factors, and physical and human capital accumulation) and hence represent *correlations* but absolutely *not causal relations* between cultural traits and economic growth. Two reasons might preclude a direct causal relation from culture to economic growth. First, culture and economic performance are so tightly linked that changes in one will work back on the other (Landes 1998, p.517). Hence, the endogenous nature of culture implies that, despite its important path-dependency (i.e. the fact that culture is an historical heritage) and the various impacts that cultural aspects can have on growth, economic development is surely associated with shifts toward values that are increasingly rational, tolerant, trusting, and participatory (Inglehart & Baker 2000). Second, numerous scholars (Todd 1983; Fischer 1989; Greif 1994; Guiso et al. 2004) have claimed that culture does not directly affect economic growth, but instead plays an indirect role through institutions, and probably as many researchers have argued that on the contrary institutions shape cultural traits (Alesina & Fuchs-Schündeln 2007; Tabellini 2010; Grosjean 2011; Nunn & Wantchekon 2011). Common sense suggest that culture and institutions are connected through a feedback relation, which is not surprising given the blurred and overlapping definitions of these two concepts.

### *1.3.3 Institutional Hypothesis*

Defining institutions

Building on North & Thomas (1973), North (1990, p.3) defines institutions as “the rules of the game in a society or, more formally the humanly devised constraints that shape human interactions. In consequence they structure incentives in human exchange, whether political, social, or economic”. But North (1994) further divides institutions into formal constraints (constitutions, rules, laws), informal constrains (norms of behavior, convention, and self-imposed codes of conduct), and their enforcement characteristics. In North’s theory, formal rules and their enforcement emanate from the polity, whereas informal norms “come from socially transmitted information and are part of the heritage that we call culture” (North 1990, p.37). The clear overlap of North (1990;1994)’s definition of institutions with Guiso et al. (2006)’s definition of culture is bridged by Acemoglu et al. (2005) who define institutions as mechanisms through which social choices are determined and implemented. These authors furthermore distinguish between economic and political institutions, and hence leave to culture the informal constraints of North. In combination with the distribution of resources[1], political institutions determine the distribution of political power across different socioeconomic groups, which in turn shape economic institutions that direct economic performance and the distribution of resources (Figure 1.11).

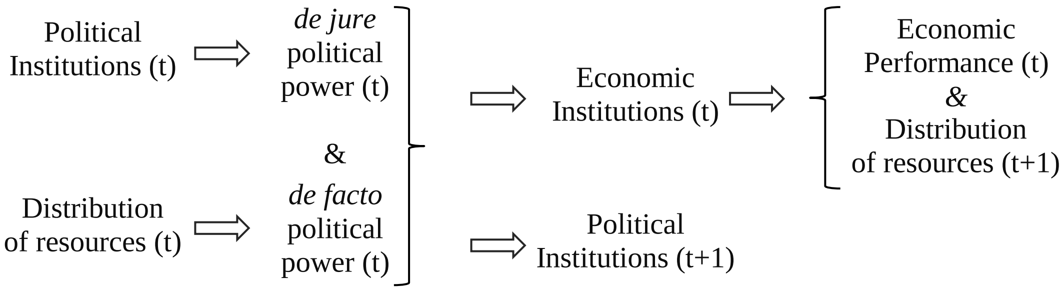


**Figure 1.11 Institutions, distribution of power and resources, and economic performances. Here, *de jure* political power refers to power that originates from the political institutions such as the form of the government (democracy vs. dictatorship or autocracy); *de facto* political power refers to the specific influence of groups due to their resources (i.e. wealth) endowment. Source: reproduced from Acemoglu et al. (2005).**

Inclusive vs. exclusive institutions: theory

The synergistic relation between economic and political institutions of Figure 1.11 is enriched by the distinction, made in Acemoglu & Robinson (2012), between *inclusive* and *exclusive* institutions. *Inclusive economic institutions* are those that allow and encourage participation by the great mass of people in economic activities, make best use of their talents and skills, and enable individuals to make the choices they wish. To be inclusive, economic institutions must feature secure private property, an unbiased system of law, and a provision of public services that provides a level playing field in which people can exchange and contract; it must also permit the entry of new businesses and allow people to choose their careers

---

[1] A word of caution is needed here: *resources* for Acemoglu et al. (2005), and more generally for neoclassical economists, absolutely do not refer to natural input resources unless stated otherwise. Conventionally, the term means economic output resources, i.e. wealth in a broad sense.

(*Ibid*., pp.74–75). *Extractive economic institutions* have opposite properties and are designed to extract incomes and wealth from one subset of society to benefit a different subset (*Ibid*., p.76). *Inclusive political institutions*, defined by these same authors as those that are sufficiently centralized and pluralistic, become *exclusive political institutions* when either of these conditions fails (*Ibid*., p.81). Economic and political institutions logically interact in strong synergy. For example, "extractive political institutions concentrate power in the hands of a narrow elite and place few constraints on the exercise of this power. Economic institutions are then often structured by this elite to extract resources from the rest of society. Extractive economic institutions thus naturally accompany extractive political institutions. In fact, they must inherently depend on extractive political institutions for their survival. Inclusive political institutions, vesting power broadly, would tend to uproot economic institutions that expropriate the resources of the many, erect entry barriers, and suppress the functioning of markets so that only a few benefit" (*Ibid*., p.81).

The central thesis of Acemoglu et al.[1] is that "economic growth and prosperity are associated with inclusive economic and political institutions, while extractive institutions typically lead to stagnation and poverty. But this implies neither that extractive institutions can never generate growth nor that all extractive institutions are created equal" (Acemoglu & Robison 2012, p.81). For these authors, "rich nations are rich largely because they managed to develop inclusive institutions at some point during the past three hundred years" (*Ibid*., p.364). On this point, it is important to understand that "inclusive economic and political institutions do not emerge by themselves. They are often the outcome of significant conflict between elites resisting economic growth and political change and those wishing to limit the economic and political power of existing elites" (*Ibid*., p.332). North (1994, pp.360–361) further emphasizes this idea that "institutions are not necessarily or even usually created to be socially efficient; rather they, or at least the formal rules, are created to serve the interests of those with the bargaining power to create new rules". That is why, Acemoglu & Robinson (2012, pp.372–376) assert that countries become failed states not because of their geography or their culture, but because of the legacy of (i) extractive economic institutions that do not create the different incentives needed for people to save, invest, and innovate, and (ii) extractive political institutions that concentrate power and wealth in the hands of those controlling the state, opening the way for public investment negligence, unrest, strife, and civil war.

### Inclusive vs. exclusive institutions: empirical evidence

Using the same data, Hall & Jones (1999), Knack & Keefer (1995), and Acemoglu et al. (2001; 2002) report a cross-country bivariate relationship between the log of GDP per capita in 1995 and a broad measure of property rights, "protection against expropriation risk", averaged over the period 1985 to 1995. Easterly & Levine (2003) assert that "measures of tropics, germs, and crops *explain* cross-country differences in economic development through their impact on institutions". In the same way, Rodrik et al. (2004) assert that the quality of

---

[1] Apart from the economic-political institutional enhancing loop theory of Acemoglu et al., other publications discussing institutional change are worth mentioning, such as: North (1981; 1990), Nelson & Winter (1982), Alston et al. (1996), Williamson (2000), Acemoglu (2005), and Easterly (2008).

institutions (property rights and *rule of law*[1]) is far more important for explaining economic growth than geography or trade. However, as in the case of culture, such quantitative studies between the quality of institutions and economic growth have two pitfalls: (i) only broad proxies are available to measure explicative variables (often a crude measure of property right protection which is one of the many aspects of the quality of institutions), (ii) econometric regressions can deliver significant *correlations* but *causal relations* cannot be formally proved, even when instrumental variables are included.

Hence, to support their theory, proponents of the *Institutional Hypothesis* rely on the narratives of natural experiments.[2] A first example is the contrast between the Democratic People's Republic of (North) Korea and the Republic of (South) Korea. As explained by Acemoglu & Robinson (2010, pp.6–7), "the Democratic People's Republic of Korea, under the dictatorship of Kim Il Sung, adopted a very centralized command economy with little role for private property. In the meantime the Republic of Korea relied on a capitalist organization of the economy, with private ownership of the means of production, and legal protection for a range of producers, especially those under the umbrella of the chaebols, the large family conglomerates that dominated the Republic of Korea's economy. Although not democratic during its early phases, the Republic of Korea's state was generally supportive of rapid development and is often credited with facilitating, or even encouraging, investment and rapid growth. Under these two highly contrasting regimes, the economies of the Democratic People's Republic of Korea and the Republic of Korea diverged. While the Republic of Korea has grown rapidly under capitalist institutions and policies, the Democratic People's Republic of Korea has experienced minimal growth since 1950, under communist institutions and policies".

A second important natural experiment of the *Institutional Hypothesis* is European colonialism. Sokoloff & Engerman (2000) develop the idea that the varying quality of institutions set up in different European colonies in the fifteenth century may have had a persistent effect on the level of development of countries once they recovered their independence. Based on this idea, Acemoglu et al. (2001) report that in colonies in which Europeans did not settle in large numbers, such as Africa, Central America, the Caribbean, and South Asia, their objective was to oppress the native population and facilitate the extraction of resources in the short run. On the contrary, in colonies where Europeans settled in large numbers, such as the United States, Canada, Australia, and New Zealand, the institutions were being developed for their own future benefits, and hence were inclusive. Acemoglu et al. (2001) further show that these different colonization strategies were in part determined by the mortality rates of settlers (they find a significant negative correlation between mortality rates of settlers and the quality of early institutions). In places where Europeans faced very high mortality rates (in particular due to malaria), they could not go and settle, and they were consequently more likely to set up extractive institutions to heavily exploit local populations and natural resources. Finally, Acemoglu et al. (2001) argue that these early institutional differences have had long-lasting effects on present income per capita distribution since they find a significant positive

---

[1] The rule of law is the legal principle that law should govern a nation, as opposed to the arbitrary decisions of individual government officials. It primarily refers to the influence and authority of law within society, particularly as a constraint upon behavior, including behavior of government officials.

[2] Theoretical frameworks have also been developed to analyze the impact of institutions on economic growth: Saint-Paul & Verdier (1993), Alesina & Rodrik (1994), Persson & Tabellini (1994), Benabou (2000), and Acemoglu & Robinson (2000a; 2006).

correlation between the quality of early institutions and present institutions on the one hand, and significant positive correlation between present quality of institutions and income per capita on the other (when controlling for latitude, climate, current disease environment, religion, natural resources, soil quality, ethnolinguistic fragmentation, and current racial composition). In another study, Acemoglu et al. (2002) furthermore report that Europeans were more likely to introduce extractive institutions in areas originally more densely populated by natives, because it was more profitable for them to exploit the indigenous population, either by having them work in plantations and mines, or by maintaining the existing system and collecting taxes and tributes. This suggests another source of variation in institutions that may have persisted to the present through delays in the onset of industrialization for countries in which extractive institutions were set up.

A third important natural experiment documented by Acemoglu et al. (2005, p.393) and Acemoglu & Robinson (2012, p.73, 197, 198, 202) concerns the underlying causes of the British success in leading the Industrial Revolution. They argue that the Atlantic trade and associated European colonialism that started in the sixteenth century brought tremendous institutional changes to Western Europe, and in particular England, that were conducive to the later industrial onset of the eighteenth century. More precisely, the growth of the Atlantic trade strengthened merchant groups by constraining the power of the European monarchies (e.g. Stuart monarchs overthrown in the Civil War and Glorious Revolution in England), and helped merchants obtain changes in institutions to protect property rights (on land and capital), which paved the way for further innovations in economic institutions. Indeed, with their newly gained property rights, English and Dutch merchant nations invested more, traded more, and spurred economic growth.

Reversing the loop and intractable endogeneity with culture

Glaeser et al. (2004) harshly criticize the institutional-change-induces-economic-growth theory of Acemoglu et al. and other *new institutionalists*. They claim that this theory fails to establish a causal link from institutions to economic growth because of conceptual problems with the measurement of institutions and the limitations of econometric techniques. According to these authors, most indicators of institutional quality used to establish that institutions cause growth are constructed to be conceptually unsuitable for that purpose. They also assert that some of the instrumental variable techniques used in the literature are flawed. Glaeser et al. (2004) suggest that human capital is a more basic source of growth than institutions are, arguing that cases exist of poor countries escaping poverty through good policies, sometimes even pursued by dictators, whose positive economic outcomes then improve political institutions. Glaeser et al. (2004) recognize that their view is clearly in line with the *Modernization theory* developed by Lipset (1960). This approach suggests that economic growth and the processes that go along with it, such as expanding education, urbanization, or the development of a middle class, determine institutional change, and not the other way around. Lipset's hypothesis was recently redeveloped by Djankov et al. (2003) and received substantial empirical support from Barro (1999) and Przeworski et al. (2000).

In addition to being hardly quantifiable, the *Institutional Hypothesis* seems trapped in an intractable endogeneity with cultural traits and economic growth aspects. Institutions seem unlikely to explain by themselves the differential growth patterns of countries, and many authors point also to the sensitivity of this hypothesis to small initial historical differences, accidental and contingent events, and a broader favorable conjuncture.

### *1.3.4 Contingency-Accident-Conjunction (CAC) Hypothesis*

Contrary to *Eurocentric* scholars who see the rise of the West as a gradual process caused by deep-rooted factors of a biogeographical, cultural, or institutional nature, the *California School of Economic History*[1] designates scholars who do not see the Great Divergence as the culmination of a long process of dynamic West vs. stagnant East, but rather as the result of different accidental events that worked in conjunction through the sixteenth to eighteenth centuries to explain the Western European (and first of all British) take-off. Of course, members of the California School do not always agree with each other in every respect, but it is fair (to my mind) to put them under a common heading that I have called the *Contingency-Accident-Conjunction Hypothesis*, or more simply the *CAC Hypothesis*.

#### Relativizing "the West and the Rest"

The first achievement of the California School has been to relativize the uniqueness of many traits of early modern European society and to rehabilitate the place of China (in particular its most proto-industrialized regions of the Yellow and Yangtze deltas) and Japan in the early modern economy (Vries 2010). Wong (1997, p.278) asserts that "China and Europe shared important similarities of preindustrial economic expansion based on Smithian dynamics. These included increased rural industries, more productive agricultures, and expanded commercial networks". In the same way, Pomeranz (2000) suggests that circa 1750 Britain, Eastern China, and Japan had many more similarities than differences in terms of capital accumulation, economic institutions (such as security of property), scale and nature of luxury demand, and even material standards of living (i.e. wages and incomes[2]). Regarding the supposed adequacy of European economic institutions to the onset of an industrial capitalism, Pomeranz (2000, p.165) argues that "when it came to matters of 'free labor' and markets in the overall economy, Europe did not stand out from China and Japan; indeed, it may have lagged behind at least China. At the very least, all three of these societies resembled each other in these matters far more than any of them resembled India, the Ottoman Empire, or southeast Asia". If Lee & Feng (1999), Goldstone (2009), Morris (2010), Hoffman (2015), and Vries (2010, 2015) are close to these views, other scholars such as Flynn & Giráldez (1997), Franck (1998), Marks (2002), Goody (2004), Hobson (2004), and Perdue (2005) are more radical and go further in asserting the backwardness of Europe and the primordial role that China played in the world economy to enable the Western European take-off. These authors argue that Western Europe "did not do anything – let alone 'modernize' – by [itself]" (Franck 1998, p.259) since it was "a peripheral, marginal player trying desperately to gain access to the sources of wealth generated in the East" (Marks 2002, p.43).

---

[1] The term was coined by Goldstone (2009) because most of the members of this approach (including Goldstone himself) worked at universities in California.

[2] On this important point, Allen et al. (2005) and Bengtsson et al. (2009) provide much quantitative evidence.

### The silver trade

The argument of the radical wing of the California School that China was the center of the early modern global economy of the fifteenth and sixteenth centuries rest on the reality of the huge import surplus of silver that flowed from the European colonies of Latin America to China in exchange for silk, ceramics, gold, copper-cash and tea exports towards Western Europe. All authors (more or less) agree on the magnitude of this silver flow and its beneficial effects for Western Europe. The difference is that some scholars (Flynn & Giráldez, Frank, Marks, Goody, Hobson, and Perdue) see in this phenomenon a clear demonstration of the economic hegemony of China, whereas for others (Wong, Pomeranz, Lee & Feng, Goldstone, Morris and Vries), it simply corresponds to the monetization of the Chinese economy and corresponds rather to a windfall for Western Europe. Hence, if for Frank (1998, p.128) "China was only able to satisfy its insatiable 'demand' for silver because it had an inexhaustible supply of exports, which were in perpetual demand elsewhere in the world economy", Pomeranz (2000, p.4) emphasizes that "the remonetization of China with silver from the fifteenth century on played a crucial part in making Spain's far-flung New World empire financially sustainable".

### Nation-state warfare, oceans sizes, and coal location

If scholars of the California School support that "the West and the Rest" were far more similar than what Eurocentrics used to say, they nonetheless recognize that important differences existed in terms of political structures. As first argued by Wong (1997) and Pomeranz (2000, p.194), the most important point is that through the fourteenth to eighteenth centuries, Europe had violent competing states (Britain, Germany, France, Spain, and Portugal) that were consequently more aggressive in their tactics of trade, whereas China was a unified, agrarian empire where elites had few institutionalized claims on the state and hence developed policies and institutions that maintained the existing social order. Hoffman (2015) developed an economic model that he confronted to political history to explain that incessant warfare among closed nation-states is not sufficient to explain the astonishing rapid growth in Europe's military sector from the Middle Ages on, which resulted in an insurmountable lead in gunpowder technology. Contrary to other parts of the world (China, Japan, India, Russia and the Ottoman Empire), Western Europeans countries not only had frequent wars, but its rulers also had lower political costs of summoning resources (through taxes), higher incentives to not use older military technologies, and few obstacles to adopting military innovations, even from opponents. For Hoffman (2015), political history explains that Western Europe acquired an insurmountable lead in gunpowder technology, which then determined which states established colonial empires or ran the slave trade, and even which economies were the first to industrialize.

In addition to these political differences, Morris (2010) highlights that Europe was lucky to have a decisive geographic advantage in reaching the New World since crossing the Atlantic was far more manageable than overcoming the huge Pacific barrier. As put by Pomeranz (2000, p.185), "the political-economic institutions of European capitalism and violent interstate competition, combined with some very lucky (for Europe) global conjunctures, made European (especially British) relations with the rest of the Atlantic world unique among core-periphery relationships".

Hence, as put by Morris (2010, p.499), “more by accident than design, western Europeans created new kinds of oceanic empires”, which, as emphasized by all proponents of the *CAC Hypothesis*, allowed the extraction of natural resources (sugar, tea, wood, fur, guano, and especially cotton) from the New World with the extensive use of slaves, and hence flooded Western European markets with new exotic products. Expanding European markets have been greatly beneficial to Western Europe and conducive to an *Industrious* Revolution in many of its constitutive states (de Vries 1994). In such Western European proto-industrial nations (in particular in Britain), wages steadily increased from the sixteenth to the eighteenth centuries, and hence incentives for labor-saving technologies were important there but inexistent in China, Japan or India where labor remained relatively cheap. Simultaneously, because proto-industry relied heavily on woodfuel, critical levels of wood scarcity, visible both in quantity shortages and price increases, were recurrent in most of Western Europe, and especially in Britain (Pomeranz 2000, pp.220–223). At these times of consequent incentives for both labor-saving and woodfuel-saving technologies in Western Europe, a fortuitous accident is emphasized by most (if not all) proponents of the *CAC Hypothesis*, it is the lucky endowment of Western European countries, and here again most notably Britain, with large and relatively accessible deposits of coal. As brightly summarized by Morris (2010, p.500), “by 1770 Britain not only had higher wages, more coal, stronger finance, and arguably more open institutions (for middle- and upper-class men, anyway) than anyone else, but–thanks to coming out on top in its wars with the Dutch and French–it also had more colonies, trade, and warships”.

Regarding the crucial importance of coal, Pomeranz (2000, p.166, 217) is highly inspiring. He calculates *ghost acreages* needed to feed and heat the British population of the nineteenth century if coal and natural resources from the colonies of the Americas had not been available. In making such calculations, he explains that without the huge consumption of coal to replace woodfuel, and the timber and calories imports of the New World, Britain and other countries of Europe would have faced an *ecological bottleneck* that would have closed the Industrial window and left Britain in the Malthusian Trap. As emphasized by Vries (2010, p.736), “before the first industrial revolution, all economies, even the most advanced ones, were Malthusian, i.e. dependent for their wealth on the quantity and quality of their land. They all faced the same constraints. They did not massively use fossil fuels. If they did use them at all, it was for heating, not as a power source. Californians endorse Wrigley’s characterization of the first industrial revolution as a process that ended this direct and full dependency on the land (Wrigley 1988; Goldstone 2002). Without it, both Britain and China, according to Pomeranz, would sooner or later have hit the Malthusian ceiling”. As successive contingent (China’s need for silver) or accidental (coal deposits, Atlantic vs. Pacific sizes) circumstances arose in a globally favorable conjuncture (rise of nation-states through warfare and early proto-industrialization), the rise of the West and relative lag of the East were surely not inevitable, but as time passed such an event clearly became increasingly probable.

This first chapter focused on the description of the four main facts of very long-term economic growth. The transition from stagnation to growth is visible in the growth rates of key variables such as per capita income, population density, fertility, and levels of industrialization and urbanization. The differential timing of the economic take-off from stagnation to sustained growth among regions of the world and the associated variations in the timing of their demographic transitions led to the phenomenon called the Great Divergence. The initial economic take-off of England from the Malthusian Epoch was associated with the Industrial Revolution that started there in 1750–60, and then spread to Western Europe and the Western Offshoots (USA, Canada, Australia, and New-Zealand) during the first part of the nineteenth century. The economic take-off of Latin America and West Asia took place toward the beginning of the twentieth century, whereas East Asia and Africa's economic take-offs were further delayed well into the twentieth century. Moreover, the human adventure seems self-organized in hierarchized-nested adaptive cycles that define both its structure and its dynamic functioning. Finally, it is clear that the improvement and diffusion of new technologies, such as general purpose technologies, is strongly connected with energy consumption.

This first chapter gave also a detailed description of the different so-called *ultimate* causes of growth that should preferably be termed *deep-rooted* causes. Several biogeographical factors have had an undeniable deep-rooted influence on the timing of the Agricultural and Industrial Revolutions. Those biogeographical factors include favourable climate conditions, size and orientation of major continental axis, length of coastline relative to mainland area, and mostly the favourable location of primary exergy flows (biomass, then water and wind) and stocks (coal, peat for the Netherland, then oil and gas). Cultural and institutional attributes are interlinked in an intractable endogeneity and seem to be more consequences than causes of economic growth and development. Finally, it is possible that some historical events (colonization, silver trade between the Americas and Europe and onward to China from 1500–1800) have generated temporary constraints that might have prevented or delayed the economic take-off of several countries.

Economists have a less narrative approach than historians. They prefer to develop mathematics-based models of economic growth that are clearly worth studying in a second chapter to understand that the true fundamental cause of economic growth that most historians and economists have ignored is useful exergy consumption, as will be explained in chapter 3.

# CHAPTER 2

# FACTS, PROXIMATE CAUSES OF ECONOMIC GROWTH, AND UGT

> *"The factors we have listed (innovations, economies of scale, education, capital accumulation, etc.) are not causes of growth; they are growth."*
>
> North & Thomas

The previous chapter focused on the different deep-rooted causes that have been proposed and studied by historians and econometricians to explain the facts of very long-term economic growth. Theoretical models developed by economists have rarely focused on these issues. Instead, economists have concentrated their efforts on designing theoretical models in order to enlighten the growth process of economies that are already in the modern regime of high sustained growth (Modern Growth Era). Such models appeal to proximate causes of growth to explain that post-WWII industrialized economies are moving along (or rather around) a balanced growth path. Contrary to the previous chapter where the facts of long-term economic growth were presented before their supposed deep-rooted causes, here I present at one and the same time both the facts about modern economic growth and the different theoretical models that economists have come up with to explain them. This approach is adopted because theoretical growth models do not prove anything *per se*. Instead, they provide a formal reflective framework to support intellectual postulates that are formulated *a priori*. This is especially evident when one considers that mainstream theoretical models are rarely simulated to reproduce historical patterns, mostly because their main variables (human capital, knowledge, skills) are not readily quantifiable. Hence, an economist would focus chiefly on the potential existence of the stable equilibrium of the model (i.e. a steady-state), the value of which is then analyzed through static comparative approaches (or econometric analyses of a reduced equation). Hence, a theoretical mathematical model of growth in which the dynamics of a given factor is central *cannot show* that this factor is a cause of growth. It can merely *formalize* the presupposed intuition that the modeler has about the causal relation of this factor and economic growth. In the final part of this chapter I discuss the contribution of the recent Unified Growth Theory (UGT), which has brought about a huge revival of interest in long-term economic growth theories.

## 2.1 FROM EXOGENOUS TO ENDOGENOUS ECONOMIC GROWTH

### *2.1.1 Physical Capital and Labor Accumulation, and Technological Change*

Political economy of the Classics and the optimal saving rate

Even if classical economists such as Smith (1776), Malthus (1798), Ricardo (1817), and Marx (1867) focused their analyses on the political aspects of the economic system and (except for Malthus) the conceptualization of *value* as a by-product of labor embodied in goods, they had some pioneering ideas about the process of economic development. While they clearly avoided the question of technological change, they all emphasized that all production processes encounter diminishing returns on a given input investment when other factors of production are held constant. Economics was then mathematicized by Jevons (1871), Menger (1871), and Walras (1874) who developed a marginal approach to the utilitarian vision of Bentham (1789) and Mill (1863) and shifted the emphasis of economics to the notions of general equilibrium and intertemporal trade-off. Several important steps had to be achieved before modern economic growth theory could properly emerge. One such step came from Pigou (1920) who suggested that people discount future utility and consequently do not save enough to provide for their later wants or, in a different context, people in every generation consume too much, leaving too little for their successors. Ramsey (1928), a young mathematics student of Keynes (1936), then tackled the issue of determining the optimal rate of saving and confirmed Pigou's conjecture that the optimal savings rate is higher than the rate chosen by myopic agents in a market economy. The idea that capital investment is a determinant of future economic growth was in place in the promising intertemporal optimal approach of Ramsey (1928), but this pioneering work was forgotten for a few decades (probably because of its high level of mathematical formalism for the time).

The Harrod-Domar model

Attention turned instead to the independent works of Harrod (1939) and Domar (1946) resulting in the so-called *Harrod-Domar model*, which is a mathematical formalization of Keynes' viewpoint.[1] In this model, output is a linear function of capital, which is an essential input. As a consequence, the marginal and average products of capital are equal and the production function exhibits constant returns to scale. Furthermore, the product of the savings rate and output equals total saving, which equals total investment; and the change in the capital stock equals investment less the depreciation of the capital stock. The main results of this model are that (i) output and capital are linearly related and have the same growth rates (in other words the capital-elasticity of output is equal to unity), (ii) the savings rate times the marginal product of capital minus the depreciation rate equals the output growth rate. Hence, in the Harrod–Domar model, economic growth arises either from an increase in the savings rate, an increase in the marginal product of capital, or a decrease in the capital depreciation rate. The implication is that policies favoring a high saving rate lead to more investment, which fosters capital accumulation and hence generates economic growth. Another conclusion of the Harrod-Domar

---

[1] Similar but less straightforward approaches were also developed independently by Feldman (1964 [1928]) and Mahalanobis (1953).

model is that an economy does not naturally achieve full employment and stable growth rates. The Harrod-Domar model was extensively used to back-up economic development policies in the 1950s, with the implication that poor countries should borrow to finance investment in capital to trigger economic growth (with the unfortunate historical consequences that this often causes repayment problems later). Eventually, two main features of the Harrod-Domar model led to its abandonment: (i) the saving rate is an exogenous parameter, (ii) there is no explicit aggregated production function since output is just linearly correlated to the capital stock through the constant output-capital ratio.

The Solow-Swan model

Solow (1956) and Swan (1956) independently worked out the second point to give the renowned Solow-Swan model in which the introduction of an aggregated production function $Y = F(K, L)$[1] allowed the inclusion of labor as an additional factor of production. The basic *Solow-Swan model* without technological change has a unique steady state with global asymptotic stability, which depends on exogenous parameters, namely the saving rate, the population growth rate, and the depreciation rate of the capital stock; but it has no sustained growth. Indeed, without technological change and starting with a sufficiently low capital-labor ratio, the basic Solow-Swan model can only generate (ever decreasing) economic growth along its transition path to the steady state, but at this stable point there is no growth in the capital-labor ratio, no more capital deepening (i.e. capital intensification), and no growth in output per capita. However, a particular setting discussed in Solow (1970) of the basic Solow-Swan model without technological change can feature sustained economic growth. This particular case is called the *AK model* because it requires the production function to be ultimately linear in the capital stock (i.e. $\alpha$ tends towards 1 in terms of the Cobb-Douglas production function). In such a case, adjustment to the steady-state would be so slow that it would mimic sustained growth and even portrayed sustained growth if the production function is instantly linear in the capital stock ($\alpha = 1$ at the initial period). However, such a setting ultimately implies that, as time goes by, the share of national income accruing to capital will increase towards 1 (if it is not equal to 1 to start with), which goes against empirical data since the capital share in national income is generally around 0.3 for developed countries (Acemoglu 2009, p.56).

The crucial role of technological change

Hence, apart from the special AK setting, the Solow-Swan model must include some technological change in order to ensure sustained economic growth. This is clearly visible in Figure 2.1a where the actual dynamics of capital and labor by themselves are unable to explain the historical course of GDP in the USA from 1900 to 2000. In this figure, capital and labor time series are aggregated in a Cobb-Douglas production function (as originally done by Solow 1957) where constant output elasticities are equal to their respective national income shares (typically 0.3 for capital and 0.7 for labor). The observed *Solow residual* is simply the difference between the actual historical US GDP and the reconstructed Cobb-Douglas US GDP, and as

---

[1] In Solow (1956), the production function had no particular form but in his following work, Solow (1957) used the famous Cobb-Douglas representation discussed later on in this section. In this particular formulation $Y = K^{\alpha}L^{1-\alpha}$, where $\alpha$ is the constant output elasticity of capital. Recall that the output elasticity of a production factor $X$ is the percentage change in output $Y$ per one percent change in input factor $X$, all other production factors remaining constant.

can be seen it is increasing over time. Hence, if the aggregate production function is to match the historical pattern more closely, a time-dependent multiplier, generally noted $A$, must be added to the production function to take into account the technological progress of the economy. In that sense, technological change has a very catch-all definition since very different production-augmenting factors are grossly aggregated in this single variable. Those include: the primary-to-final and final-to-useful energy conversion efficiency (if energy is considered as an input factor), the division and organization of labor, the broader organization and efficiency of markets, the skill improvements of laborers, the contribution of information and communication technologies, but also the beneficial effects of inclusive institutions (which, for example, protect private property rights and consequently incentivize innovation and R&D).

If one considers as in Figure 2.1b that technology improves the efficiency with which both capital and labor inputs are used then the technology is said to be *Hicks-neutral* and the time-dependent multiplier $A$ is called Total Factor Productivity (TFP). With such a definition of technology, the aggregated production function takes the general form $\tilde{F}(K,L,A) = AF(K,L)$. In the US case, Figure 2.1b shows that such TFP broadly matched an exponential function with a 1.6% annual growth rate.

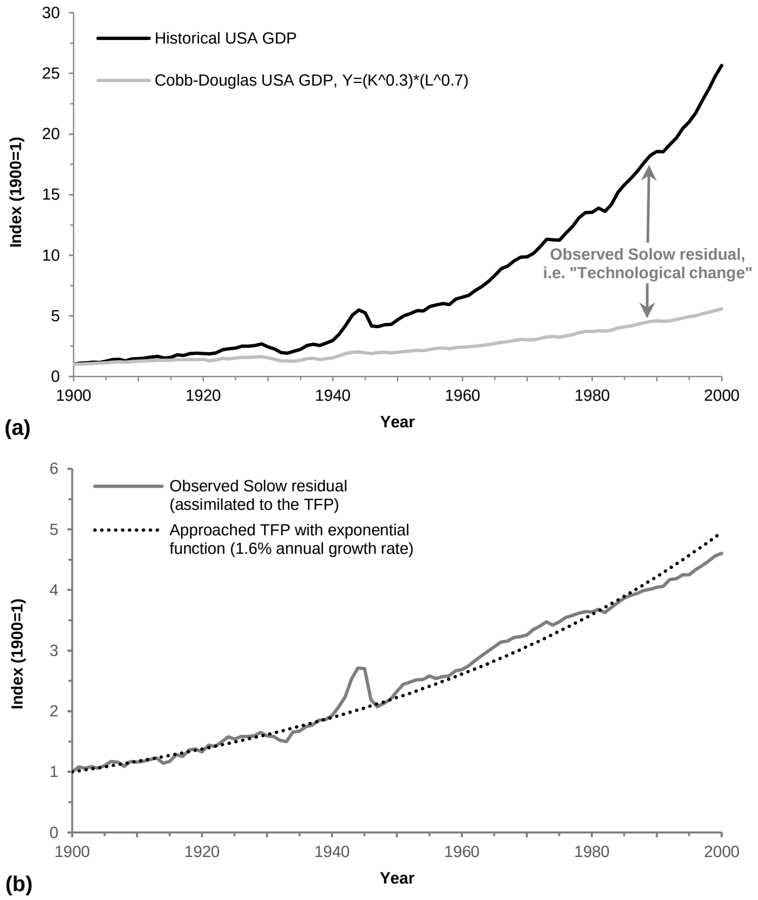


**Figure 2.1 US GDP and Total Factor Productivity (TFP), 1900–2000.**
**(a) Historical US GDP vs. reproduced US GDP with historical capital and labor aggregated in a Cobb-Douglas function. (b) Observed vs. approached (with exponential function) Solow residual. Data source: Warr et al. (2010).**

On the other hand, one can only consider capital-augmenting or *Solow-neutral* technological change in the form $\tilde{F}(K, L, A) = F(AK, L)$. And a third alternative is to have labor-augmenting or *Harrod-neutral* technological change in the form $\tilde{F}(K, L, A) = F(K, AL)$. This last option is often preferred by economists because the first theorem of Uzawa (1961) states that technological change must be asymptotically Harrod-neutral for balanced growth to be possible[1]. It will be recalled that a balanced growth path (BGP) is defined by a constant rate of growth for the per capita output, and constant interest rates, capital-output ratio, and national income shares for factors of production.

As suggested in Figure 2.1 technological change seems to explain the largest part of the economic growth process in the Solow-Swan model. But contrary to what one might think, it is difficult to be more precise than that because the literature on growth estimates is vast and different approaches such as TFP accounting, cross-country regressions, and national calibrations are used and debated (Solow 1957; Denison 1967; 1985; Barro 1991; Barro & Sala-I-Martin 1992; 2004; Jorgenson et al. 1987; Jorgenson 1995). At the risk of oversimplification, the following order of magnitude can be noted: for a 1% output growth in the Solow-Swan framework, the relative contributions of labor, capital, and technology accumulation are respectively 15%, 25%, and 60%. The main problem is of course that in such a framework the technological change that accounts for the lion's share of economic growth is completely exogenous and factors that induce some firms and societies to invent and adopt better technologies remain to be elucidated (models dealing with such issues are reviewed in the next section). Regarding capital accumulation, the Solow-Swan model delivers the idea that countries with higher saving rates, a lower depreciation rate, and lower population growth will have higher capital-labor ratios and hence will be richer. But here again those features are determined by purely exogenous parameters.

Rediscovering Ramsey (1928) and following the work of von Neumann (1945), Cass (1965) and Koopmans (1965) explicitly modeled the intertemporal utility optimization of a representative household in the economy. This led to an endogenization of the saving rate of the household within a neoclassical economic growth setting, a framework also called the *Ramsey-Cass-Koopmans model*, or sometimes more simply the neoclassical growth model. The interesting thing about this model is that the discount rate affects the rate of capital accumulation. Broadly speaking, a lower discount rate implies greater patience, thus greater savings and hence greater capital accumulation and economic growth.[2] As in the Solow-Swan model, technological progress is indispensable to have sustained economic growth in the Ramsey-Cass-Koopmans model, and here again balanced growth is only possible if all technological change is asymptotically purely labor-augmenting (Harrod-neutral), and if furthermore the elasticity of intertemporal substitution (which determines the speed of adjustment to the steady state) tends towards a constant. What is clearly visible here is that the Ramsey-Cass-Koopmans model does not provide new insight into the causes of economic growth but only clarifies the nature of the economic decisions.

---

[1] The first theorem of Uzawa and its corollary do not state that technological change has to be labor-augmenting all the time, but rather that it ought to be labor-augmenting after some time $T$ along the balanced growth path (Acemoglu 2009, p.62).

[2] Another important result is that the form of the utility function affects the transitional dynamics but has no impact on steady states (Acemoglu 2009, p. 301).

### *2.1.2 Human Capital and Knowledge Spillovers*

Human capital in exogenous growth models

Before turning to the first class of endogenous growth models, we must discuss the inclusion of human capital in the Solow-Swan, AK, and Ramsey-Cass-Koopmans models. The notion of human capital, generally noted $H$, was elaborated in the seminal work of Becker (1965), and Mincer (1974). In this conventional[1] view, human capital represents the stock of skills, education, competencies, and other productivity-enhancing characteristics embedded in labor; in other words, it represents the efficiency units of labor (Acemoglu 2009, p.85). Mankiw et al. (1992) were the first to provide a Solow-Swan model augmented with human capital (and purely labor-augmenting technological change). From a theoretical point of view this model has the same comparative statics as the original Solow-Swan model, so it still needs increasing technological progress to have sustained economic growth. However, when performing cross-country regression analyses, Mankiw et al. (1992) found that differences in factors endowments (human and physical capital, and routine labor) explains 78% of the observed differences in GDP level across countries, thus considerably lowering the relative role of technological change compared to previous estimates. The methodology employed by Mankiw et al. (1992) was criticized by Klenow & Rodriguez (1997) and Hall & Jones (1999) who proposed an updated value of around 60% (hence placing greater emphasis on the importance of technological change differences across countries as a source of growth differential). The work of Barro (1991) and Barro & Sala-i-Martin (1992; 2004) is instrumental in completing this picture and specifying the notion of *convergence* among countries and regions since the Great Divergence. These authors find no convergence among countries of the entire world, which contradicts the basic Slow-Swan and Ramsey-Cass-Koopmans models because countries with initially lower capital-labor ratios should have had higher growth rates and converged towards initially richer countries. However, when focusing on clusters of countries which have much more similar levels of education, institutions, policies, and initial conditions than the world as a whole, it is possible to find some *conditional convergence*, meaning convergence after controlling for measures of education and government policies. For example, Barro (1991) and Barro & Sala-i-Martin (1992; 2004) find that the income gap between countries that have the same human capital endowment has been narrowing over the postwar period on average at about 2 percent per year. In the same way, these authors show that after controlling for measures of education and government policies, poor countries tend to grow faster than rich one. As explained by Jones (1997), another important finding of this literature is that one can interpret the variation in growth rates around the world as reflecting how far countries are from their steady state positions. For example, Korea and Japan grew rapidly in the 1980s–1990s because their steady state positions in the income distribution were much higher than their actual positions. Venezuela grew slowly because the reverse was true. On these mixed results, Acemoglu (2009, p.105) suggests that even if a complete consensus is impossible, it is fair to say that differences in physical and human capital stocks and growth cannot by themselves explain differences

---

[1] I use this term to make a distinction between the traditional Becker-Mincer approach to human capital and the less accepted but complementary Nelson-Phelps-Schultz perspective described later on in this section.

across countries in terms of GDP level and growth, so that technology plays an essential role in such dynamics.

### Other aspects of human capital

Moreover, if human capital can easily be incorporated into the Ramsey-Cass-Koopmans model or in the AK model (see Acemoglu 2009, p.367 and 393 respectively), other aspects of human capital are worth mentioning. First, if the literature on schooling typically finds that one more year of schooling increases earnings by about 6–10%,[1] the *Ben-Porath model* (1967) shows that there is also on-the-job human capital accumulation after school years. This model also suggests that in countries with high schooling investments, one can also expect higher levels of on-the-job investments in human capital, which would tend to prove that there is a systematic mis-measurement of the amount or quality of human capital across societies. Second, empirical evidence, from Griliches (1969) and Krussell et al. (2000) notably, suggests that physical and human capital are complementary so that productivity can be lower than its best potential in case of imperfect labor markets in which, as modeled by Acemoglu (1996), factor prices do not necessarily reflect marginal products. Third, human capital can also have technological externalities as first emphasized by Jacobs (1969), and later by Lucas (1988) and Azariadis & Drazen (1990). Such technological externalities must be understood as the local consequences of human capital concentration which can affect competitive markets and prices. Fourth, according to Schultz (1964) and Nelson & Phelps (1966), the major role of human capital might not be to increase productivity in existing tasks, but to enable workers to cope with change, disruptions, and the implementation of new technologies. In that sense the Nelson-Phelps-Schultz view of human capital is different from the more conventional Becker-Mincer definition. At this point, the modern growth process is still obscure because the dynamics of the proximate causes of growth (physical and human capital accumulation and technological change) are determined by exogenous parameters (population growth rate, depreciation rate, technological growth rate, saving rate or, discount rate and elasticity of intertemporal substitution in Ramsey-like settings) that have no clear links with the deep-rooted (biogeographical, cultural, institutional, lucky) causes of growth presented in Section 1.3.

### First models of endogenous growth

The first model of quasi-endogenous economic growth came from Romer (1986) who regards knowledge accumulation as a by-product of physical capital accumulation. In this model, the production function aggregates purely physical capital with routine labor and the technology is exclusively labor-augmenting. The key concept is then to consider that although firms take the technological level $A$ as given, this stock of knowledge is a linear function of the physical capital stock $K$. In this sense, technology is the result of spillovers from physical capital and it evolves endogenously for the economy as a whole. This assumption appeals to the concept of *learning-by-doing*, whereby greater investment in a given sector increases the experience (of workers and managers) in the productive process and hence makes the production process itself more productive (Arrow 1962). A very important point to notice is

---

[1] This is the general agreement found for example by Card (1999), but Acemoglu & Angrist (2000) find much more lower external returns to schooling of 1–2% (which are moreover statistically insignificant), a result confirmed by Duflo (2004) and Ciccone & Peri (2006).

that the linear relation between technology and physical capital developed by Romer (1986) implies increasing returns to scale in the overall production process, whereas the different models studied so far (i.e. the different forms of the Solow-Swan, AK, and Ramsey-Cass-Koopmans models) have constant returns to scale. While arguably crude, Romer (1986)'s formalization also captures the idea that knowledge is a *nonrival* good, meaning that, once a particular technology has been discovered, many firms can make use of this technology without precluding its use by others.[1] Another important paper came from Lucas (1988) who constructs a model with a similar structure except that technology evolves as a linear function of human and not physical capital (i.e. Romer has physical capital externalities, whereas Lucas has human capital externalities). A third important model of the bourgeoning endogenous growth literature of the 1990s is the two-sector AK model of Rebelo (1991). In this model the consumption good uses capital and labor in a Cobb-Douglas function with a Hicks-neutral technology, while the investment good only consumes capital as an input. This formulation has the advantage of presenting sustained economic growth with constant factor shares in national income, contrary to the original one-sector AK model (or its revised forms with human capital and/or endogenous savings) in which the capital income share is ultimately equal to unity.

In the models presented so far, sustained economic growth is either the result of exogenous technological change or a by-product of endogenous knowledge spillovers from physical or human capital accumulation. The following section investigates models in which economic growth results from technological progress itself as a consequence of purposeful investments by firms and individuals.

### *2.1.3 Expanding Variety and Schumpeterian Models, and Directed Technological Change*

#### Economic growth with expanding input variety

Inspired by the Dixit-Stiglitz (1977) model of industry equilibrium, Romer (1987) was the first to provide an economic growth model based on expanding input varieties. In this kind of model (also called the lab-equipment model of growth), different *intermediate machine inputs* are aggregated with labor in the final good sector, and the economic growth potential is a function of the number $N$ of these different intermediate input goods. Furthermore, the number of machines developed by the intermediate sector depends on the level of research and development (R&D) expenditure. With free entry for firms into the R&D sector, greater spending is rewarded with a perpetual monopolistic position on the blueprint or idea that is invented to produce the machine, which leads to the increasing invention of new machines (i.e. hence the name *expanding variety model*) and consequently to an increase in the final output good, so to economic growth. In another paper, Romer (1990) separates the rival component of knowledge, i.e. human capital $H$, from the nonrival component, i.e. the number of designs or machines $A$. In doing so he introduces knowledge spillovers from human capital $H$ toward the technological component $A$, a phenomenon also known as the *standing on giant's shoulders effect*. Basically, contrary to Romer (1987) where R&D is fueled by scarce physical capital expenditure, in Romer (1990) R&D is a function of potentially infinite human capital. This

---

[1] On the contrary knowledge can be an *excludable* good if patent rights and copyrights are implemented to protect the results of research and development (R&D).

specification led Romer (1990, p.99) to conclude about his model that its most interesting positive implication “is that an economy with a larger total stock of human capital will experience faster growth”, and “that low levels of human capital may help explain why growth is not observed in underdeveloped economies that are closed and why a less developed economy with a very large population can still benefit from economic integration with the rest of the world”. Another interesting point raised by Romer is the implication of his model for the phenomenon of escape from the Malthusian Trap described in Section 1.1.1. Indeed, although he refers to stagnation as the prehistoric and not pre-industrial period, Romer (1990, p.96), stresses that “if the total level of human capital is too small, stagnation may arise”, which “is reminiscent of the explanation for the absence of growth in prehistoric time that is offered by some historians and anthropologists: civilization, and hence growth, could not begin until human capital could be spared from the production of goods for immediate consumption”. And quite logically, regarding the phenomenon of Great Divergence presented in Section 1.1.2, Romer (1990, p.96) asserts that his analysis also “offers one possible way to explain the wide variation in growth rates observed among countries and the fact that in some countries growth in income per capita has been close to zero”. These observations are important to understand the path chosen by Galor (2011) to develop his Unified Growth Theory presented in Section 2.3.

### Economic growth with Schumpeterian innovation

Before turning to this point, it is important to note that models of expanding machine variety may not provide a good description of innovation dynamics in practice because they do not capture its qualitative aspect. Indeed, in addition to the increasing number of products (which is sometimes referred to as horizontal innovation), Schumpeter (1942) theorized the concept of *creative destruction* by which economic growth is mainly driven by the innovation of new machines and products replacing old ones, and hence possibly new firms replacing incumbents. That is why models discussed in the quality ladder realm are also called *Schumpeterian growth models*. The first of these models is probably[1] Segerstrom et al. (1990), but Grossman & Helpman (1991b) and Aghion & Howitt (1992) are more renowned.[2] Basically, contrary to expanding variety models in which economic growth was determined by the *increasing number of machines* (or blueprints/designs), in Schumpeterian models the engine of growth is the process of innovations that leads to the *increasing quality improvements* of a fixed number of machines. Formally, the idea is to think that “there is a quality ladder for each machine variety, and each innovation takes the machine quality up by one rung on this ladder” (Acemoglu 2009, p.459). A crucial assumption is that the different qualities of the same machine are perfect substitutes and that in equilibrium only the leading-edge (i.e. highest quality) version of each machine type is used. This aspect is at the center of the process of creative destruction because when a higher quality machine is invented it will replace the previous vintage of the same machine which becomes useless and is consequently destroyed.

---

[1] I use this formulation because the publication dates of peer-reviewed papers do not reflect the fact that earlier working paper versions can rapidly influence researchers. For example, Grossman & Helpman (1991b) cite Aghion & Howitt (1990) which is the working paper version of Aghion & Howitt (1992).

[2] Aghion & Howitt (1998) provide an excellent survey of many Schumpeterian models of economic growth and numerous extensions regarding employment, step-by-step vs. cumulative innovations, and so on.

Another very important aspect of Schumpeterian growth models is that only new entrant firms undertake R&D. This is logical because the incumbent has weaker incentives to innovate since it would replace its own machine for which it enjoys the benefit of a perpetual patent. In such models, it is shown that there is no transitional dynamics and there exists a unique BGP in which the average quality of machines, output, and consumption grow at the same rate. An interesting difference is worth noting between the two-sectors models of Segerstrom et al. (1990) and Grossman & Helpman (1991b) on the one hand, and the one-sector model of Aghion & Howitt (1992) on the other. In the latter, only one sector experiences quality improvements rather than a continuum of machine varieties, and the innovation possibilities frontier uses a scarce factor–labor–as in the model of knowledge spillovers of Romer (1990). This implies that in the only one sector experiencing technological change, growth only occurs at finite intervals and consequently the growth rate of the economy is *uneven* in its nature. Hence, in Aghion & Howitt (1992) the economy has a constant output for a precise interval of time and experiences a burst of growth when a new machine is invented. Whether this pattern of uneven growth provides a better approximation to reality than the continuous growth of Segerstrom et al. (1990) and Grossman & Helpman (1991b) is open to debate. Indeed, Acemoglu (2009, p.470) remarks that “while modern capitalist economies do not grow at constant rates, they also do not have as jagged a growth performance as that implied by” Aghion & Howitt (1992).

Some criticisms and the direction of technological change

Two important points must be mentioned about the different endogenous models previously reviewed. First, these models (except Romer 1990) include a *scale effect* so that a higher population leads to a higher growth rate. As noticed by Jones (1995), the consequence of this characteristic is that if the population is growing in such frameworks, then the economy would not admit a balanced growth path but would lead to an exploding path and to infinite utility for the representative household (Acemoglu 2009, p.401, 446). Second, while the Solow-Swan and Ramsey-Cass-Koopmans models have difficulties in generating very large income differences across countries, the models presented above suffer from the opposite problem. Indeed, with exogenous economic growth models (Solow-Swan and Ramsey-Cass-Koopmans settings), even quite large differences in cross-country distortions (e.g. eightfold differences in effective tax rates) do not generate large income per capita differences in the steady-state, which contradicts the reality of the Great Divergence presented in Section 1.1.2. On the other hand, with endogenous economic growth models (expanding input variety and Schumpeterian models), even small differences in policies, technological opportunities, or other characteristics of societies, lead to permanent differences in long-run growth rates. This outcome means that these models are better equipped to explain the phenomenon of Great Divergence, but at the same time they predict an ever-expanding income distribution across countries whereas data suggest relative stability since the Second World War (Acemoglu 2009, p.403).

Furthermore, in all the endogenous models presented previously, technological change (by knowledge spillovers, increasing input varieties, or increasing input quality) increases the aggregate productivity of the economy, but in practice technological change is often directed or *biased* towards one kind of agents or another. Indeed, historical evidence supports the idea that during the nineteenth century technological change was *unskilled-biased* (i.e. favoring unskilled laborers), but in the early twentieth century this phenomenon was reversed so that technological change has been increasingly *skilled-biased* during the last one hundred years in developed countries.[1] These facts received early attention from the literature on *induced innovation* started by Hicks (1932, pp.124–125) who argues that: “A change in the relative prices of the factors of production is itself a spur to invention, and to invention of a particular kind—directed to economizing the use of a factor which has become relatively expensive”. The concept of “innovation possibilities frontier” of Kennedy (1964), and the emphasis of Habakkuk (1962) on the search for labor-saving inventions in the USA and the UK during the nineteenth century because of labor scarcity and increasing wages are in the same line of thought but lack any mathematical micro-foundations. Such formalism on directed technological change was recently developed within expanding input variety or Schumpeterian frameworks, with noticeable references such as Acemoglu (1998; 2002a; 2003; 2007), Duranton (2004), Jones (2005) and Caselli & Coleman (2006).[2] These models are important evidence that within the neoclassical realm: (i), technological change is rather logically labor-augmenting (as opposed to capital-augmenting) because of gross complementarity between labor and capital; and (ii), the direction of technological change towards skilled and unskilled labor depends on their relative scarcity and their elasticity of substitution.

While the three previous sections have highlighted how the proximate causes of economic growth have been inserted into mathematical representation, the next section focuses on other important mechanisms that were not included in these models. A first shortcoming shared by all of them is to consider that each country is an island (i.e. a closed economy) that does not interact with the rest of the world. This is detrimental for two reasons. First, many countries not only generate technological change from their own R&D but also benefit from the advances in the world technology frontier. Second, international trade of financial capital and commodities has an influence on economic growth, so this interaction must be analyzed. Furthermore, models related to structural change, namely the sectorial evolution of the economy, and market failures are worth studying because they have been important in paving the way towards the Unified Growth Theory presented at the end of this chapter.

## 2.2 TECHNOLOGY DIFFUSION, TRADE, AND STRUCTURAL CHANGE

### *2.2.1 Technology Diffusion*

#### The role of human capital to cope with the world technology frontier

As already stressed, one important criticism of the Solow-Swan and the Ramsey-Cass-Koopmans models regards their inability to generate quantitatively large differences in cross-country income per capita, and many economists relate this to the incapacity of these models to provide an explanation for technological differences across countries. Based on the essay of Gerschenkron (1962), Acemoglu (2009, pp.613–618) provides an augmented Ramsey-Cass-

---

[1] For the quantitative evidence of these two successive phenomena accompanying the economic development of Western European countries and the Western Offshoots, see James & Skinner (1985), Mokyr (1990), and Goldin & Katz (1998) for the unskilled bias of the nineteenth and early twentieth century, and Autor et al. (1998), Katz & Autor (1999), and Acemoglu (2002b) for the accelerating skilled bias in the twentieth century.

[2] Furthermore, the link between directed technological change and unemployment is analyzed in Blanchard (1997), while Thoenig & Verdier (2003) focus on the influence of international trade.

Koopmans model in which the world economy consists of $J$ countries, indexed by $j = 1, \dots, J$, each with access to a technology $A_j$ that is lower than the world technology frontier $A$ encapsulating the maximal knowledge that any country can have. The interesting feature of this model is to consider that each country absorbs the world technology at some exogenous technology absorption rate $\sigma_j$, and also improves its technology thanks to local R&D at the exogenous speed $\lambda_j$. Then, it is assumed that these parameters depend on the specific human capital stocks of each country, with the absorption rate $\sigma_j$ being linked to the Nelson-Phelps-Schultz approach of human capital, and $\lambda_j$ corresponding to the more traditional Becker-Mincer view. Indeed, it will be recalled that the Nelson-Phelps-Schultz view of human capital emphasizes the role of human capital in facilitating the adoption of new technologies and adapting to changing environments, which is exactly the role of $\sigma_j$ measuring the adoption of the existing technologies and adaptation of existing blueprints to the specific conditions prevailing in the country, so that they can be used with the other technologies and practices already in place. On the other hand, the exogenous speed of technological accumulation $\lambda_j$ could be identified as the more traditional Becker-Mincer propensity of human capital to determine the effectiveness of the R&D process. Since in this model $\sigma_j$ multiplies $\left(A - A_j\right)$, an important implication is that countries that are relatively backward (in the sense of having a low $A_j$ compared to the world technology frontier $A$) tend to grow faster because they have a higher technological gap to resorb (i.e. more room to catch-up). In particular in the steady state, this force pulling backward economies toward the technology frontier is powerful enough to ensure that all countries grow at the same rate, which is the exogenous growth rate of the world technology frontier. Thus, differences in saving rates, absorptions rates $\sigma_j$, and the specific technological speed of convergence $\lambda_j$ translate into *level and not growth rate differences* across countries in this model.

Institutional barriers preventing technology diffusion

An alternative interpretation of the absorption rates $\sigma_j$ is to link them to differences in institutional barriers to technology adoption (property rights, taxes, or other policy features). This is the option chosen by Parente & Prescott (1994) who propose a model of technology diffusion within an expanding input variety setting. In this model investment affects technology absorption and countries differ in terms of the barriers that they place in the path of firms in this process. Similarly, Howitt (2000) introduces a source of differences in the cost of technology adoption across countries within a Schumpeterian growth framework. As in the simpler model of Acemoglu (2009, pp.613–618) previously studied, all countries grow at the same rate in these models and differences in the cost of technology adoption determine differences in the level of per capita income across countries. However, the advantage of these last two models (Parente & Prescott 1994 and Howitt 2000) is that the rate of growth of the world technology frontier, and hence of economic growth, is endogenous. These perspectives are interesting but how exactly institutions affect technology is left as a black box. An intuitive answer to such a question is given by Acemoglu et al. (2007). In a Schumpeterian setting, this paper illustrates how contractual difficulties can affect the relationship between producers and suppliers and thus change the profitability of technology adoption, which leads to differences in technology adoption and productivity patterns across countries.

### Inadequacy of the world technology frontier with developing countries needs

In the absence of institutional barriers, it is possible that technological differences and income gaps remain because the world technology frontier is inappropriate to the needs of specific countries. If that is the case, importing the most advanced frontier technologies may not guarantee a convergence of the productivity of all countries. And indeed since OECD countries are both the producers of much new know-how and the largest markets for new technologies, it is logical to expect that new technologies are optimized for the conditions and the needs of these countries and not for developing ones. This view can be clarified following Atkinson & Stiglitz (1969) who model technological change as shifting isoquants (increasing productivity) at a given capital-labor ratio. Hence, if new technologies are developed for high capital-intensive production processes in OECD countries (say ICT-assisted tractors that increase labor productivity in the agricultural sector), they may be of little use to labor-abundant, less-developed economies which have far lower capital-output ratios (and use less advanced tractors or even oxen). This point is developed in the context of a Solow-type growth model by Basu & Weil (1998) whereas Acemoglu & Zilibotti (2001) discuss more broadly the implications of the mismatch between technologies developed in advanced economies and the skills of the workforce of the less-developed countries.

### Product cycle and associated technology transfer

Technology diffusion is logically linked to international trade in commodities. Vernon (1966) was the first to suggest this point through his concept of *product cycle*, whereby the life cycle of (high-income and usually labor-saving) new products is divided into three successive steps: New Product, Maturing Product, and Standardized Product.[1] A New Product is typically developed in a high-tech country before expanding its international market through exports towards middle-income countries. Once a foreign market is more secured (in less technologically advanced countries), it becomes worthwhile producing a Maturing Product locally. Finally, when the product is fully standardized it can be produced in low-income low-technology countries. In some situations, the product moves so far away from its point of origin that it becomes an item that is imported by its original country of invention. This pattern goes hand-in-hand with technological diffusion from highly technologically advanced economies to less technologically developed countries. Krugman (1979) was the first to model the associated process of international product cycle and technology diffusion. An interesting implication of this model regards international protection of intellectual property rights (IPR). It can be shown that stronger international IPR protection invariably increases the income gap between the high-income high-tech country and the low-income low-tech country.

---

[1] A more precise product lifecycle would now be divided into four stages: Introduction, Growth, Maturity, and Decline.

### *2.2.2 International Trade*

International trade of financial capital and equalization of effective capital-labor ratios

When witnessing the important cross-country difference in income per capita that has existed since the Great Divergence, one might wonder why this gap has not been narrowed by a flow of financial capital from rich to poor countries. In a globalized economy if the rates of return on capital differ across countries as a consequence of different capital-labor ratios, we would expect capital to flow towards areas with lower capital-labor ratios where rates of return on capital are higher, i.e. towards poorer countries (Acemoglu 2009, p.648). As a consequence of such financial flows, capital-labor ratios of countries would equalize and economies would converge in terms of income per capita. But Feldstein & Horioka (1980) pointed out early on that there is actually much less net flow of capital from countries with high saving rates toward those with lower saving rates than a theory of perfect international capital markets would suggest. One way to explain this point would be to simply remark that the international capital market is surely not frictionless and that additionally sovereign risk probably prevents such flows.[1] But in fact, Lucas (1990) in a Solow-Swan model with human capital and Acemoglu (2009, pp.649–653) in a Ramsey-Cass-Koopmans setting, show that even with perfect international capital markets, if the productivities (i.e. technology levels) of countries are different, capital flows equalize *effective* capital-labor ratios (i.e. capital-labor ratios times technological level) but this does not imply the equalization of capital-labor ratios. Hence, in such models there is no reason to expect financial capital to flow from rich to poor countries. As a result, in international borrowing and lending models, there are only transitional dynamics for the aggregated world economy, but no transitional dynamics separately for each country.[2]

International trade of commodities, specialization, and equalization of effective factors prices

Apart from international trade in financial assets, international trade in commodities is perhaps even more important for the economic growth process. Within the neoclassical paradigm, international trade is usually studied within a *Heckscher-Ohlin model* (1967),[3] which reformulates the idea of *comparative advantage* first enunciated by Ricardo (1817). Both approaches share identical hypotheses regarding the perfect mobility of productive factors within the different sectors of a given country, the immobility of factors between countries, and the international free trade in final goods[4]. In the basic Ricardian model, only labor is considered as a production factor and trade is ultimately motivated by differences in labor productivity, meaning that countries have different technological levels that determine their

---

[1] The paper by Obstfeld & Taylor (2003) contains a survey of the literature on why capital does not flow from rich to poor countries. Other interesting references on this topic are Kehoe & Perri (2002), and Matsuyama (2004).

[2] As pointed out by Acemoglu (2009, p.653), "international capital flows ensure that each country has the same effective capital-labor ratio; thus dynamics resulting from slow capital accumulation are removed. The corollary therefore implies that any theory emphasizing the role of transitional dynamics in explaining the evolution of cross-country income differences must implicitly limit the extent or the speed of international capital flow."

[3] The original essay of Ohlin was published in 1933. Although Ohlin wrote the book alone, his doctoral thesis director Heckscher was credited as co-developer of the model which was formalized mathematically in the second edition of 1967.

[4] This last assumption is probably the most extreme since trading internationally involves costs and many analyses of international trade incorporate the physical costs of transportation and tariffs. The most important implication of this assumption is that the prices of traded commodities—final goods in the original *Heckscher-Ohlin* model, intermediate goods in augmented versions such as Trefler (1993) and Ventura (1997)—in all countries are equal to their world prices, determined by the world supply and demand for these commodities.

comparative advantage. In this approach, without technological differences between countries, all nations would become autarkic at various stages of growth with no reason to trade with each other. On the contrary, in the basic Heckscher-Ohlin model, technologies are assumed to be identical in all countries but within countries each productive sector has different capital-, labor-, and land- intensities (i.e. sectors within countries have different technologies). As a result, it is the relative endowments of the factors of production (capital, labor, and land stocks) that determine a country's comparative advantage because each country specializes in the production of goods for which the required factors of production are relatively abundant locally. When there is a sufficient difference in factor endowments across countries to ensure that they arbitrage relative differences in factor costs and consequently perform international trade, countries are said to be in the *cone of diversification*. The first and most important result of the basic Heckscher-Ohlin model is that a country will export goods that use its abundant factors intensively, and import goods that use its scarce factors intensively. In other words, a capital-abundant country (such as the USA) will export capital-intensive goods while a relatively labor-abundant country (such as India) will export labor-intensive goods. This result was challenged early on by Leontief (1953) who noticed that the United States, despite having a relative abundance of capital, tended to export labor-intensive goods and import capital-intensive goods. This inconsistency of the Heckscher-Ohlin model with empirical data was named the *Leontief paradox* after that study. The second most important result of the basic Heckscher-Ohlin model is that the factor prices (rates of return on capital and land, and labor wages) of all countries ultimately converge and are equal at the steady-state. Though important, this second result is not supported by compelling evidence in empirical data. Neither the rental return on capital and land, nor the wage rates seem to consistently converge between trading partners at different levels of development. Trefler (1993) proposed an augmented Heckscher-Ohlin model allowing for exogenous differences in technologies between countries and free trade in intermediate inputs instead of final goods. These amendments, especially the inclusion of differences in labor productivity across countries, explain the so-called Leontief Paradox and in fact show that there is no such enigma: factor-augmenting international technology differences imply that endowments must be adjusted to reflect international productivity differences. Trefler (1993) recognizes that Leontief (1953) himself had raised this point and that it had been mistakenly forgotten by scholars who wanted to undermine the Heckscher-Ohlin model. As a result, in such a setting there is convergence of *effective* factor price equalization (meaning factor prices corrected for countries intrinsic differences of productivity), a phenomenon Trefler (1993) called *conditional factor price equalization*, which is far more in line with empirical data. For example, it is necessary to take into account that British labor productivity is only two-thirds of U.S. labor productivity to explain why British wages are about two-thirds of U.S. wages. The third most important result of the Heckscher–Ohlin model (known as the Stolper–Samuelson theorem) is that a rise in the relative price of a good will lead to a rise in the return to that factor which is used most intensively in the production of the good, and conversely, to a fall in the return to the other factor. The same result is displayed by the *Specific-Factors model* originally developed by Jones (1971) and Samuelson (1971) as an augmented Ricardian model with two goods that require specific immobile factors of production (capital and land respectively) and unspecific mobile labor. An important corollary of this result for discussions regarding the impact of internationalization on within-country inequalities is that unskilled workers producing traded goods in high-skill countries (USA, Western Europe) will be worse off and capital owner will be better off as international trade increases, because, relative to the world market in the good they produce, an unskilled first world production-line worker is a less abundant factor of production than capital.

### International trade and the stability of the world income distribution

Regarding economic growth, an important feature of the basic and augmented Heckscher-Ohlin models is that there is no transitional dynamics for each country, but only a general transitional dynamics for the world economy which ultimately converges towards a stable steady-state. The stability of this global steady-state equilibrium results from the *integration* of the world economy. An even more interesting result of these models comes from the fact that while the world economy has a standard neoclassical production function, each country generally faces an AK technology and thus can accumulate as much capital as it wishes without running into diminishing returns (as long as the country remains small, which is always a valid hypothesis in a medium-run perspective). As a consequence Ventura (1997) proposes that the augmented Heckscher-Ohlin model with conditional factor price equalization can easily rationalize the growth miracles that countries can sometimes display for a few decades. In particular, this mechanism can explain that between the 1960s and 1990s, thanks to their greater openness to international trade, the East Asian Tigers (Hong Kong, Singapore, South Korea, and Taiwan) accumulated capital more rapidly than many other developing countries without experiencing diminishing returns and consequently witnessed sustained growth at far higher rates than the world average. Hence, similarly to the financial borrowing and lending models described previously, commodity trade in Heckscher-Ohlin settings emphasizes the potential pitfalls of using closed-economy growth models for the analysis of output and capital dynamics across countries and regions (Acemoglu 2009, p.662).

The role of international trade in ensuring a stable world income distribution is perhaps even clearer in the work of Acemoglu & Ventura (2002). In this Ricardian model, all countries have different technologies and consequently specialize in the production of one of the $N$ intermediate inputs. This assumption, commonly called the *Armington preference*, ensures that while each country is small in import markets, it has *market power* in the goods that it supplies to the world. As a consequence, when a country accumulates capital faster than the rest of the world and thus increases the supply of its exports relative to the supplies of other countries' exports, the price of its export goods declines relative to other countries' goods and hence it will face worse *terms of trade* (the price of its exports relative to its imports). This negative terms-of-trade effect reduces its income and its rate of return on capital, which slows down capital accumulation. As a consequence, the world economy and in fact all national economies, move toward a unique stable steady-state world where all countries grow at the same exogenous rate. In other words, international trade, together with terms-of-trade effects, lead to a stable world income distribution in this model (Acemoglu 2009, p.670), a feature of the real world that other models rarely present. One might see an inconsistency between the Heckscher-Ohlin model of Ventura (1997) discussed above and the present Ricardian setting of Acemoglu & Ventura (2002). Indeed, in the latter higher than average capital accumulation cannot last because of the worsening terms-of-trade effect, whereas in the former fast capital accumulation without diminishing returns was indispensable to explain how certain economies can grow rapidly for extended periods. This conflict is resolved if one assumes that an early developing country is specializing and thus can accumulate capital without diminishing returns and worsening terms of trade, whereas a more advanced economy produces more differentiated goods and is consequently better described by a situation of diminishing returns on capital and terms-of-trade effect.

Regarding the question of the impact of trade openness on economic growth more specifically, the literature is slightly mixed but more in favor of a positive effect. Frankel and Romer (1999) logically find that international trade in a model of expanding input variety implies that trade encourages technological change and increases the growth rate of the world economy. On the contrary in endogenous models with learning-by-doing, Young (1991) and Matsuyama (1992) find that international trade is detrimental to economic growth in developing countries. On the empirical side, Dollar (1992) and Sachs & Warner (1995) find a positive correlation, hence not a causal one, between openness to international trade and economic growth.

### *2.2.3 Structural Change and Market Failures*

#### Economic growth vs. economic development

The models studied so far provide a good approximation of the behavior of relatively developed societies since they focus on either balanced growth paths (BGPs) or transitional dynamics leading to BGPs. But countries in earlier stages of development have many salient features that are hardly compatible with BGPs, which implies that the different models studied previously in this chapter cannot describe them adequately. This highlights the need, avoided so far, to distinguish between the notions of economic growth and economic development. *Economic growth*, as stressed in the introduction to this thesis, formally represents the annual rate of growth of the macroeconomic output (generally the Gross Domestic Product, or GDP). *Economic development* is a related notion that more largely encompasses the structural changes (sectoral shifts in production and employment, level of financial development, demographic transition, urbanization, migration, etc.) and the efficiency implications of such transformations that particularly affect economies in the early stages of development. In particular, one might expect structural change to be accompanied by a process that involves the organization of production becoming more efficient and the economy moving from the interior of the aggregate production possibilities set toward its notional frontier. In addition to the study of current developing countries, it seems clear that models of economic development are needed to explain the (first two) facts of very long-term economic growth and the relation between technological change and energy consumption presented in Chapter 1. This inadaptability of the modern growth theory to describe the dynamics of countries from stagnation to growth has motivated the research for a Unified Growth Theory (UGT). Before presenting the first advances of UGT, some prerequisite studies are worth examining.

#### Demand-side and supply-side structural change

Changes in the composition of employment and production between agricultural, manufacturing, and service sectors are a significant component of the *nonbalanced* or *uneven* growth associated with economic development. Caselli & Coleman (2001) and Kongsamut et al. (2001) deliver tractable frameworks in which Engel’s Law augmented with Stone-Geary preferences implies that, as they become richer, households spend a smaller fraction of their budget on agricultural goods and a larger fraction on manufactured goods and services. In such a setting, an equilibrium with fully balanced growth is impossible and the different sectors grow at different rates with a differential allocation of labor and capital. This feature generates a *demand-side* structural change that reproduces the changes in the sectoral composition of employment and production. A more flexible and richer approach is to allow for *hierarchies of needs* in consumption, whereby households consume different goods in a particular sequence (typically food needs to be consumed before textiles, and textiles before electronics, and so on). Matsuyama (2002) and Buera & Kaboski (2009) use such an approach to obtain a similar demand-side structural change. Baumol (1967) proposes an alternative *supply-side* view to why growth may be nonbalanced. He suggests that nonbalanced growth is in fact a general feature of the growth process because different sectors naturally have different levels of technological change and hence grow at different rates. Following this line of thought, Ngai & Pissarides (2007) and Acemoglu & Guerrieri (2008) developed models in which factor proportion differences across sectors combined with capital deepening lead to nonbalanced economic growth. By imposing weak restrictions on the functional forms of the different sectors, both models can at the same time have uneven growth of sectors and ensure aggregate balanced growth of the economy (i.e. aggregate ratios of the economy are constant). Finally, some economic historians (Mokyr 1993; 2009; Overton 1996) have argued that the high agricultural productivity of England in the eighteenth century facilitated the possibility of an industrial revolution because England could afford more easily than other countries to shift part of its labor force to industrial activities. Matsuyama (1992) proposes a model which formalizes this intuition.

Lack of financial development

Less-developed economies may be within their production possibility set envelope because of severe market failures, to the extent of being stuck in development traps. The lack of financial development is often suggested as a possible source of drag on economic growth (Goldsmith 1969; Shaw 1973). Greenwood & Jovanovic (1990) and Bensivenga & Smith (1991) provide theoretical models in which an intermediation sector (i.e. banks) permits the economy to reduce the fraction of its savings held in the form of unproductive liquid assets. As a result, the composition of savings is altered in a way that is favorable to capital accumulation and hence that is conducive to economic growth. In a different theoretical setting, Obstfeld (1994) and Acemoglu & Zilibotti (1997) study the relationship between financial development on the one hand, and risk diversification and economic growth on the other. Working with a panel of cross-country and time-series observations, Loayza & Rancière (2006) find that a positive long-run relationship between financial intermediation and output growth co-exists with a mostly negative short-run relationship. The conclusions of Benhabib & Spiegel (2000) are similar but more nuanced since they find that the results of their econometric regressions are sensitive to the inclusion of country fixed effects, which may indicate that the financial development indicators serve as proxies for broader country characteristics. Arcand et al. (2012) suggest that that finance starts having a negative effect on output growth when credit to the private sector reaches 100% of GDP. The authors assert that their results are not driven by output volatility, banking crises, low institutional quality, or by differences in bank regulation and supervision. Hence, empirical results seem less unanimous than theoretical models in supporting the positive effect of financial development on economic growth (see also Cournède et al. 2015; Cecchetti & Kharroubi 2015).

### The dual economy and the allocation of labor

Another potentially important market failure concerns the migration of labor from rural to urban areas and hence the optimal allocation of labor between productive sectors (agriculture and manufacturing). Lewis (1954) developed the concept of *dual economy* in the pioneering work of this literature. According to this notion, less-developed countries consist of a traditional rural sector with surplus labor available to a modern urban sector. If the supply of this surplus labor is unlimited the modern sector can expand for some time without the need to raise wages which results in higher capital returns reinvested in capital accumulation. In turn, the increase in the capital stock leads to a further demand for labor, and this positive feedback becomes self-sustaining, leading the economy to a growth take-off. However, several barriers can limit the migration of the surplus labor and lead to a development trap. In a theoretical model Banerjee & Newman (1998) represent such a barrier as the advantage of rural communities over urban centers in the reduction of moral hazard problems in credit relations (i.e. borrowing and lending is easier in the traditional sector in spite of higher productivity in the modern sector). In a quite similar approach, Acemoglu & Zilibotti (1999) focus on the accumulation of information through task repetition which implies lower *agency costs* for the well-established traditional rural sector compared to the emergent modern urban sector. In a quite different way, the barrier preventing the migration of labor from the traditional rural sector toward the modern urban sector could be the result of the inadequacy between technologies and skills as already developed in Section 2.2.1 in the context of international technology diffusion. Indeed, if the skills of labor are not improved through education and the accumulation of human capital, skilled laborers will be too scarce for the modern sector to expand and the economy might remain in its dual configuration (Acemoglu 2009, p.743).

One important aspect of economic development has been deliberately discarded so far. It concerns the demographic transition accompanying the transition from stagnation to sustained growth. Including demographics in theoretical models requires the use of a particular kind of analytical framework called *overlapping generation (OLG) models*, in which the unique (normative) representative household with an infinite planning horizon is replaced by overlapping households with limited lifetimes. OLG settings are useful for studying the role of national debt and social security on economic growth within the neoclassical framework. De La Croix & Michel (2002) provide a comprehensive analysis of such intergenerational issues within the modern regime of sustained economic growth. An OLG setting is also required to incorporate the phenomenon of demographic transition associated with the process of transition from stagnation to growth assessed in the Unified Growth Theory.

## 2.3 UNIFIED GROWTH THEORY (UGT)

The term Unified Growth Theory (UGT) was coined by Galor (2005) to categorize theories of economic growth that capture the entire growth process in a single framework of analysis. With several co-authors, Galor has developed a Unified Growth Theory that captures in a single analytical framework: (i) the epoch of Malthusian stagnation that characterized most of human history; (ii) the escape from the Malthusian trap and the associated spike in the growth rates of income per capita and population; (iii) the emergence of human capital formation in the process of development; (iv) the onset of the demographic transition; (v) the contemporary era of sustained economic growth; and (vi) the divergence in income per capita across countries.

### *2.3.1 Benchmark Model and Central Mechanisms*

<u>Description</u>

The central UGT model is developed in Galor & Weil (2000) and it is presented in more detail in Galor (2011, p.140–178). The theory is based on the interactions of several building blocks: the Malthusian elements, the engines of technological progress, the origins of human capital formation, and the triggers of the demographic transition. The first block, the Malthusian Epoch, is captured in UGT by three elements: (i) the production process is characterized by decreasing returns on labor due to the limited availability of land; (ii) parents generate utility from consumption and having children but child rearing is time intensive; and (iii) individuals are subject to a subsistence consumption constraint.[1] Concerning the engines of technological progress, Galor & Weil (2000) build on Boserup (1965) and Kremer (1993) to postulate that, in the Malthusian Epoch, it is the scale of the population that affects the rate of technological change via its effect on (i) the supply of innovative ideas; (ii) the demand for innovations; (iii) the rate of technological diffusion; (iv) the degree of specialization in the production process and thus the extent of learning by doing; and (v) the scope for trade and thus the extent of technological imitation and adoption. However, as advancement of the technological frontier becomes increasingly complex in the later stages of development, human capital formation becomes the prime engine of technological change, and educated individuals are more likely to advance this frontier. Furthermore, Galor & Weil (2000) suppose that it is changes in the economic environment due to technological progress that triggered the rise in industrial demand for human capital formation since educated individuals have a comparative advantage in adapting to the new technological environment.[2] Formally, the demand for education is a function of the *rate of change* of technology and not its *level*, otherwise changes in technology could only be skill-biased and we have already seen that on the contrary technological change was unskilled-biased during the Industrial Revolution (see Section 2.1.3, p.48).

Galor’s UGT follows Becker (1981) in postulating that the rise in demand for human capital triggered the decline in fertility in the course of the demographic transition. According

---

[1] The physiological foundations of the subsistence consumption constraint and the resulting Malthusian equilibrium are analyzed by Dalgaard & Strulik (2015).

[2] Clearly, in the Galor-Weil benchmark model of UGT, the role of human capital is defined in the Nelson-Phelps-Schultz sense and not in the Becker-Mincer alternative approach (see Section 2.1.2 for explanations of these different conceptions of human capital).

to this view, (adult) individuals generate utility from the quantity and the quality of their children as well as from their own consumption (of a representative composite good). They choose the optimal number of children and their quality in the face of a constraint on the total amount of time that can be devoted to labor market activities and child rearing. While a rise in direct parental income (due to a rise in demand for human capital) would generate conflicting income and substitution effects and would not necessarily trigger a decline in fertility, the effect of a rise in demand for human capital in potential future earnings of a child generates a pure substitution effect. Hence, it induces parents to substitute quality for quantity of children and thus reduces fertility (Galor 2011, p.149). As already stressed, the phenomenon of demographic transition is essential for the onset of the state of sustained economic growth because it brings a reversal of the dominant mechanisms prevailing in the Malthusian Epoch. More precisely, the demographic transition enhances labor productivity and the growth process via three channels. First, the decline in population growth reduces the dilution of the stocks of capital, land, and infrastructures, which increases the amount of resources per capita. Second, the reduction in fertility rates enhances human capital accumulation through the reallocation of resources from the quantity of children toward their quality. Third, the decline in fertility rates affects the age distribution of the population, temporarily increasing the fraction of the labor force in the whole population, and thus mechanically increasing the labor productivity per capita.

It is worth highlighting that the increase in the pace of technological change has two opposing effects on the evolution of population. On the one hand, it eases households' budget constraints, allowing the allocation of more resources for raising children. On the other hand, it induces a reallocation of these additional resources toward child quality. In the Post-Malthusian Regime, due to the limited demand for human capital, the first effect dominates, and the rise in real income permits households to increase their family size as well as the quality of each child. Then, the interaction between investment in human capital and technological change generates a virtuous circle: human capital formation prompts faster technological change, further raising demand for human capital, inducing further investments in child quality, and ultimately triggering the demographic transition. The offsetting effect of population growth on the growth rate of income per capita is eliminated, and the interaction between human capital accumulation and technological change permits a transition to a state of sustained economic growth.

Dynamics

Hence, the core mechanisms of the Galor-Weil benchmark model are the following: (i) a scale effect of the population on technological change, (ii) the role of human capital to cope with the changing technological environment, and (iii) the effect of the rise in demand for human capital on potential future earnings of a child to direct the quality vs. quantity arbitrage of adults. Formally, the education level of workers in period $t+1$ (as determined by parents in period $t$) depends only on the technological change $g_{t+1}$ (which is the rate of growth of the technology level $A$) expected between $t$ and $t+1$, while for a given population size $L$ technological change $g_{t+1}$ between $t$ and $t+1$ depends only on the level of education in period $t$. Hence, the evolution of technological change $g_t$, and education $e_t$, for a given population size $L$, is characterized by the sequence $\{g_t, e_t; L\}_{t=0}^{\infty}$that satisfies in every period $t$ the equations $g_{t+1} = g(e_t; L)$ and $e_{t+1} = e(g_{t+1})$. Figure 2.2 depicts the three qualitative configurations of this dynamical system.

Consider an economy in the early stages of development. The population $L^s$ is relatively small, and the implied slow rate of technological change $g_t$ does not provide an incentive to invest in the education $e_t$ of children. As depicted in Figure 2.2a in such conditions the dynamical system is characterized by a globally stable steady-state equilibrium $(0, g^s(L^s))$, where $g^s(L^s)$ increases slowly with population size, while the level of education remains zero. This steady-state equilibrium corresponds to a globally stable conditional Malthusian steady-state equilibrium. For a constant small population and for a given rate of technological change, effective resources per capita and level of education are constant, and output per capita is therefore constant as well. Moreover, shocks to population or resources are resolved in a classic Malthusian fashion (increase in income per capita is rapidly offset by an equivalent increase in population). The inherent Malthusian interaction between population size and the level of technology gradually increases population and technological change, which generates an upward shift in the curve $g(e_t; L)$. Logically, as the population increases from small to moderate $(L^s \rightarrow L^m)$, the steady-state equilibrium $(0, g^s(L^s))$ shifts upward as well towards $(0, g^m(L^m))$, reflecting small increments in technological change while the level of education remains constant at zero.

Ultimately, the upward shift is sufficient for the rate of technological change to exceed the threshold level $\hat{g}$ above which investment in human capital is beneficial. As depicted in Figure 2.2b, in such a situation with a moderate population size $L^m$, the dynamical system of education and technology is characterized by three history-dependent steady-state equilibria: $(e^u(L^m), g^u(L^m))$ is an interior unstable equilibrium, whereas $(0, g^s(L^m))$ and $(e^h(L^m), g^h(L^m))$ correspond to multiple locally stable conditional steady-state equilibria. As in the initial regime of Figure 2.2a, $(0, g^s(L^m))$ is a Malthusian steady state characterized by constant resource per capita, slow technological change, and no education. By contrast $(e^h(L^m), g^h(L^m))$ is a modern growth steady state characterized by a growing income per capita, moderate technological change, and a high level of education. Given the initial conditions, in the absence of large shocks (i.e. if the economy starts in the Malthusian state), the economy remains in the vicinity of the low steady-state equilibrium $(0, g^s(L^m))$, where education is still zero but the rate of technological change is moderately higher than before. As the rate of technological change continues to rise in reaction to the increasing and now large population $L^l$, the $g(e_t; L^l)$ curve further shifts upward and ultimately, as depicted in Figure 2.2c, the dynamical system undergoes another qualitative change. The Malthusian steady-state vanishes, and the conditional dynamical system is characterized by a unique, globally stable, modern steady-state equilibrium $(e^h(L^l), g^h(L^l))$, with high levels of education, technological change, and output. In the Modern Growth Era, resources per capita rise as technological change outpaces population growth. As the size of the population increases, its effect on the rate of technological change declines asymptotically to zero. From then on the $g(e_t; L^l)$ locus no longer shifts upward and the growth rates of output and income per capita, technological change, population and education level converge to constants.

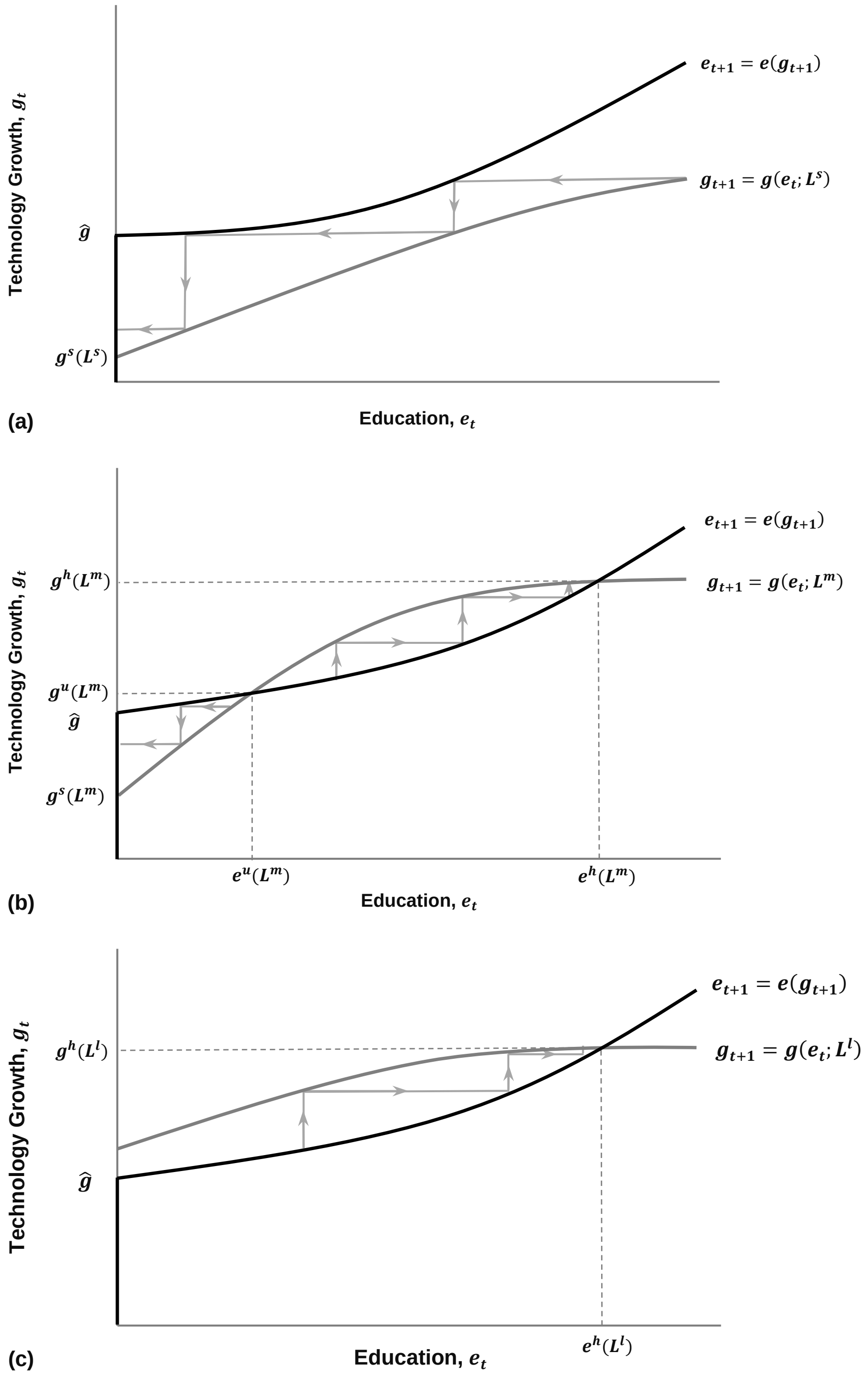


**Figure 2.2 The evolution of technology and education in the Galor-Weil model.**
**With (a) small, (b) moderate, and (c) large populations. Source: reproduced from Galor &Weil (2000).**

### *2.3.2 Comparative Development and Complementary Mechanisms*

Description

As described by Galor (2011, p.182–183), “Unified Growth Theory contains two theoretical black boxes: the effect of population size and level of human capital on the rate of technological progress and the effect of the rate of technological progress on human capital formation”. But “variation in these characteristics across countries should account for variations in economic development around the globe". In any country $i$, for given levels of population and human capital $\left(e^i, L^i\right)$, the rate of technological change $g_t^i$ is governed by a number of country-specific characteristics that are associated to the different deep-rooted causes of growth (biogeography, culture, institutions, luck) identified in Section 1.3. For instance, the propensity of a country to trade, reflecting its geographical characteristics as well as its trade policy, may foster technological diffusion across nations (Hausmann et al. 2007). Natural resources abundance may have an adverse effect on technological adoption in the absence of bad institutions (Van Der Ploeg 2011). The composition of cultural and religious groups in a society and their attitude towards knowledge creation may affect incentives to innovate and the proliferation rate of innovations (see Section 1.3.2). The composition of interests groups in society may affect incentives to block or promote technological innovation (Acemoglu et al. 2005). The level of protection of intellectual property rights may have an ambiguous effect on technological change, reflecting the trade-off between the positive effect on the incentive to innovate and its adverse effect on the proliferation of existing knowledge (Diwan & Rodrik 1991). The degree of credit-market imperfections and inequality in a society may affect human capital formation, entrepreneurial activities, and technological advancements (Aghion et al. 2005).

On the other hand, human capital formation, hence investment in education $e_t^i$, also depends on a number of country specific factors. Geographical attributes and their effects on the health environments influence the extent of underinvestment in human capital (Gallup et al. 1999). International trade enhances specialization of industrial economies in the production of skill-intensive goods, which may thus provide a further inducement for human capital formation (Galor & Mountford 2006; 2008). The composition of religious groups in society and their attitudes towards literacy and education may affect the incentive of individuals to invest in human capital formation (see Section 1.3.2). The prevalence of human capital-promoting institutions or policies (e.g. the availability, accessibility, and quality of public education and child labor regulations) affects the extent of human capital formation (Hanushek & Woessmann 2008). It may partly reflect the distribution of ownership over factors of production and landed aristocracy (Acemoglu & Robinson 2000b; Galor et al. 2009). The degree of credit market imperfections, and more precisely the (in)ability of individuals to finance the cost of education and the foregone earnings associated with schooling influence their ability to implement a desirable level of investment in education (Fernandez & Rogerson 1996).

Dynamics

Thus, in a comparative development perspective, the Galor-Weil benchmark model presented in the previous section can be augmented by the introduction of two additional country-specific vectors $\Omega^i$ and $\Psi^i$ that encompass all country-specific factors that respectively affect technological change and human capital formation. A clarification that is worth noting for the remainder of the discussion concerns the division of the country-specific vector $\Psi^i$ that affects human capital formation between the degree of preferences of adult individuals for child quality, captured by the preference parameter $\mu^i$ in the household's utility function, and $\phi^i$ which encompasses all other country-specific factors affecting human capital formation (cost of education, availability and efficiency of education in different segments of society, etc.), so that in summary $\Psi_t^i \equiv \left[\phi_t^i, \mu_t^i\right]$. Then, the evolution of technology and education in country $i$ for a given population size $L^i$ and country-specific attributes $\Omega^i$ and $\Psi^i$ is characterized by the sequence $\left\{g_t^i, e_t^i; L^i, \Omega^{\mathrm{i}}, \Psi^{\mathrm{i}}\right\}_{t=0}^{\infty}$such that in every period $t$: $g_{t+1}^i = g\left(e_t^i; L^i, \Omega^i\right)$, and $e_{t+1}^i = e\left(g_{t+1}^i;\ \Psi^i\right)$. The UGT framework can be used to study how country-specific characteristics that have affected the rate of technological progress and the formation of human capital have contributed to the differential pace of transition from stagnation to growth leading to the phenomenon of Great Divergence. For example, consider two economies, $A$ and $B$, that are identical in all respects, except for country-specific characteristics that are conducive to technological progress. In particular, the countries are identical in the characteristics that contribute to human capital formation (i.e. $\Psi^A = \Psi^B = \Psi$). Thus, for any given rate of technological progress, $g_{t+1}$, human capital formation is equal in the two economies, namely $e_{t+1}^A = e_{t+1}^B = e(g_{t+1}; \Psi)$; and the threshold of the rate of technological change above which parental investment in human capital is beneficial, $\hat{g}(\Psi)$, is also equal in the two countries. Suppose further that country-specific characteristics that are conducive to technological progress, $\Omega^i$ with $i \in (A, B)$, are more prevalent in country $B$. Hence, as depicted in Figure 2.3a, for any given level of population $L$, and human capital $e_t$, the rate of technological change is higher in country $B$ than country $A$, that is $g_{t+1}^B = g(e_t; L, \Omega^B) > g_{t+1}^A = g(e_t; L, \Omega^A)$. In the Malthusian Epoch, while income per capita in the two economies may be equal, for a given level of (small) population, the steady-state equilibrium level of education and technology in country $B$, $(0, g^s(L^s, \Omega^B))$, is higher than in country $A$, $(0, g^s(L^s, \Omega^A))$. The inherent Malthusian interaction between population size and level of technology in each country gradually increases the population size ($L^s \rightarrow L^m \rightarrow L^l$) and the rate of technological change. Thus, the potential demands for human capital also increase, generating an up-ward shift in the curves $g(e_t; L, \Omega^A)$ and $g(e_t; L, \Omega^B)$. Ultimately, the rate of technological change in country $B$ increases sufficiently, and as depicted in Figure 2.3b, crosses the threshold level $\hat{g}(\Psi)$ above which parental investment in human capital is beneficial. The conditional Malthusian steady-state equilibrium vanishes in country $B$, and the economy takes off to a conditional sustained-growth steady-state equilibrium $(e^h(L^l,\ \Omega^B, \Psi), g^h(L^l, \Omega^B, \Psi))$. In contrast, country $A$ experiences a later take-off. Moreover, if the country-specific characteristics of the two economies do not converge in the long-run, country $B$ will have a higher steady-state equilibrium. As performed above with country-specific characteristics that have affected the rate of technological change, the UGT framework is also helpful in analyzing how country-specific characteristics affecting human capital formation have contributed to the differential pace of the transition from stagnation to growth and differential development around the world. For the sake of brevity such an analysis is not presented here but can be found in Galor (2011, p.191–207).

In addition to its adaptability to comparative development, Galor (with different co-authors) has supplemented the UGT framework with several additional mechanisms related to human evolution. Two of these theoretical mechanisms (Galor & Moav 2002; Galor & Michalopoulos 2012) suggest that the struggle for existence that has characterized most of human history stimulated a process of natural selection that generated an evolutionary advantage for individuals whose characteristics (respectively for each paper: preference for offspring quality, and entrepreneurial spirit) were favorable to economic growth. In a quite different approach, Ashraf & Galor (2013b) develop the idea that there is an optimal mix of genetic diversity that is conducive to higher economic growth.

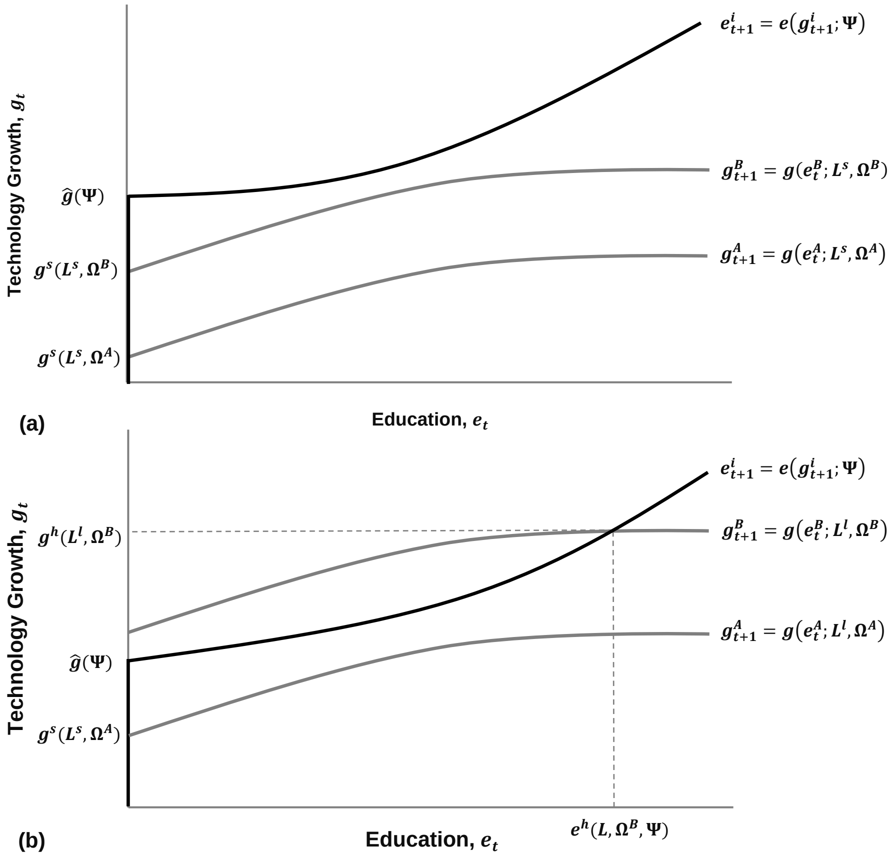


**Figure 2.3 Technological change differences and comparative development in UGT. With (a) small and (b) large population Source: reproduced from Galor (2011).**

### Darwinian selection of higher valuation for offspring quality

Hence, Galor & Moav (2002) suggest that during the epoch of Malthusian stagnation, traits of higher valuation for offspring quality generated an evolutionary advantage and their representation in the population gradually increased. Hence, this selection process is incorporated in the UGT framework through variable $\mu_t^i$ (recall that $\Psi_t^i \equiv [\phi_t^i, \mu_t^i]$) and its effect on investment in human capital stimulated technological progress and initiated the enhancing feedback (previously exposed in Section 2.3.1) between human capital formation and technological change, which ultimately brought about the demographic transition and the state of sustained economic growth. The explanation of Galor & Moav (2002) is that the division of labor and trade relationships among individuals and communities have been ever-increasing since the Neolithic Revolution, enhancing the complexity of human interactions and raising returns on human capital. In such a context, individuals born to parents with traits of (moderately) higher valuation of offspring's quality generated higher income and, in the Malthusian Epoch, when child rearing was positively affected by aggregated resources, such higher-income higher-quality oriented households generated a larger number of offspring. As the fraction of individuals with high valuation for child quality continued to increase, technological change intensified, further raising the rate of return on human capital and inducing households to reallocate these increased resources to child quality. Ultimately, further increases in the rate of technological change induced a more widespread investment in human capital along with a reduction in fertility rates, generating a demographic transition in which the rate of population growth declined along with an increase in the average level of education. The positive feedback between technological progress and the level of education reinforced the growth process, setting the stage for a transition to a state of sustained economic growth. Galor & Moav (2002) therefore suggest that the transition from stagnation to growth is an inevitable by-product of the interaction between the composition of the population (between quality and quantity child oriented individuals) and the rate of technological progress in the Malthusian Epoch. Failed take-off attempts were due to a shortage of individuals of the quality type in the population. However, for a given composition of population, these authors suggest that the timing of the transition may differ significantly across countries and regions due to variations in biogeographical, cultural and institutional factors, as well as historical accidents (trade patterns, colonial status, etc.).

Darwinian selection of entrepreneurship

Similarly to Galor & Moav (2002) who focus on the valuation of offspring quality, Galor & Michalopoulos (2012) argue that a Darwinian evolution of entrepreneurial spirit played a significant role in the process of economic development and the dynamics of inequality within and across societies. The study advances the hypothesis that entrepreneurial spirit evolved non-monotonically in the course of human history. In the early stages of development, risk-tolerant, growth-promoting traits generated an evolutionary advantage and their increased representation accelerated the pace of technological progress and the process of economic development. In the mature stages of development, however, risk-averse traits gained an evolutionary advantage and individuals characterized by entrepreneurial traits had an evolutionary disadvantage. As a consequence, natural selection ultimately diminishes the growth potential of advanced economies and contributes to convergence in economic growth across countries. Furthermore, Galor & Michalopoulos (2012) advance that in the least advanced economies, selection of growth-promoting traits has been delayed by detrimental factors of geographical, cultural or institutional nature leading to the persistence of poverty. Historical variations of such factors across countries affected the pace of the evolutionary process of growth-promoting traits selection and thus contributed to the contemporary sustained differences in productivity and income per capita across countries. Clark (2007) proposes a very similar view associating Darwinian natural selection of the fittest (in his view the richest) endowed with growth-compatible characteristics (entrepreneurial and hard-working spirits) to explain the phenomenon of Great Divergence. Being an economic historian, his approach rests on comparative qualitative analyses of historical facts and data and not on mathematically formalized models as Galor & Moav (2002) and Galor & Michalopoulos (2012).

Long-lasting effect of genetic diversity

All these *Evolutionary Economics* theories are applicable to either social or genetic intergenerational transmission of traits[1] but a more recent set of publications by Galor and colleagues explores the influences of the long-lasting genetic composition of populations on the comparative economic performance of societies. Ashraf & Galor (2013b) assert that in the course of the exodus of *Homo sapiens* out of Africa, variation in migratory distance from the cradle of mankind to various settlements around the globe negatively affected genetic diversity,[2] which had a direct, persistent, hump-shaped effect on the pattern of comparative economic development, reflecting the trade-off between the beneficial and the detrimental effects of diversity on productivity. The hypothesis rests upon two fundamental building blocks. First, migratory distance from the cradle of humankind in East Africa had an adverse effect on the degree of genetic diversity within ancient indigenous settlements across the globe. Following the prevailing hypothesis, commonly known as the *serial founder effect*, it is postulated that, in the course of human expansion over planet Earth, as subgroups of the populations of parental colonies left to establish new settlements further away, they carried with them only a subset of the overall genetic diversity of their parental colonies. Using data on genetic diversity from the 53 ethnic groups across the globe that constitute the Human Genome Diversity Cell Line Panel,[3] Ashraf & Galor (2013b) show that migratory distance from East Africa has an adverse effect on genetic diversity so that genetic diversity is higher for natives of Africa, lower for natives of Asia, Oceania and South America, and intermediate for natives of Europe. Second, to the authors' mind, genetic diversity is both negatively associated with the extent of cooperative behavior, as it raises the likelihood of disarray and mistrust (Ashraf & Galor 2013a), and positively associated with innovative activity, as measured by the intensity of scientific knowledge creation (see econometric regressions at the end of Ashraf & Galor 2013b). Hence, the degree of diversity in a society may provide a wider spectrum of traits that are complementary to the implementation of advanced technological paradigms (possibility of expanding the society's production frontier), but it may also reduce trust, cooperation and hence the efficiency of the production process. In support of their theory, Ashraf & Galor (2013b) obtain a hump-shaped relationship (i.e. an inversed U curve) when population density in 1500 CE or the level of income per capita in 2000 CE[4] is plotted as a function of genetic diversity. These hump-shaped impacts seem robust to controls for continental fixed effects, ethnic fractionalization, various measures of institutional quality (i.e., social infrastructure, an index gauging the extent of democracy, and constraints on the power of chief executives), legal

---

[1] The interaction between cultural and genetic evolution has been intensively explored since the seminal work of Cavalli-Sforza & Feldman (1981), more recent references include Richerson & Boyd (2005), Schaller (2010), and Ember et al. (2012).

[2] Population geneticists typically measure the extent of diversity in genetic material across individuals within a given population (such as an ethnic group) using an index called *expected heterozygosity*. Like most other measures of diversity, this index may be interpreted simply as the probability that two individuals, selected randomly from the relevant population, differ genetically from one another with respect to a given spectrum of traits. Specifically, the *expected heterozygosity measure* for a given population is constructed by geneticists using sample data on allelic frequencies; i.e., the frequency with which a gene variant or allele (e.g., the brown versus blue variant for the eye color gene) occurs in the population sample. Given allelic frequencies for a particular gene or DNA locus, it is possible to compute a gene-specific heterozygosity statistic (i.e., the probability that two randomly selected individuals differ with respect to the gene in question), which when averaged over multiple genes or DNA loci yields the overall expected heterozygosity for the relevant population (Ashraf & Galor, 2013, p.4).

[3] The Human Genome Diversity Cell Line Panel has been compiled by the Human Genome Diversity Project (HGDP) in collaboration with the Centre d'Etudes du Polymorphisme Humain (CEPH).

[4] Most advanced development is marked by higher population density in the Malthusian Epoch, whereas higher income per capita or GDP per capita is a better definition of higher development in the Modern Growth Era.

origins, major religion shares, the share of the population of European descent, years of schooling, disease environments, and other geographical factors that have received attention in the empirical literature on cross-country comparative development. While Galor (2011, p.219–221) shows how this theory can be embedded within the UGT framework, additional co-authored research is currently being pursued to support this theory (Arbatli et al. 2015; Ashraf et al. 2015).

Genoeconomics in the turmoil

As could have been expected, this recent work on the relationship between human selection and evolution, the genetic composition of populations, and the comparative economic performance of societies has triggered an extensive and vibrant debate.[1] For the sake of consistency and brevity, the different points raised in these criticisms and their counter-arguments are not detailed here but references are given to supplement this literature review. Among dozens of (generally positive) reviews, the evolutionary theory of Clark (2007) has come in for particularly vigorous criticism from four referees (McCloskey 2008; Voth 2008; Grantham 2008; Persson 2008) to which Clark gave a published response (Clark 2008). The work of Ashraf & Galor (2013b) has received wider attention and press coverage. In particular, a team of anthropologists (Guedes et al. 2013) made a very harsh criticism, concluding that "Ashraf and Galor's (2013b) paper is based on a fundamental scientific misunderstanding, bad data, poor methodology, and an uncritical theoretical framework". Ashraf and Galor have responded to these criticisms in an open letter available online (Ashraf & Galor 2014).

This thesis is not the right place to further investigate this controversial debate. However, it is important to review some criticisms of the more crucial mechanisms of UGT in order to close this chapter on conventional economic growth theories.

### *2.3.3 Criticisms and Alternative Mechanisms*

Scale effect and historical population sizes

In the current UGT the transition from stagnation to sustained growth is an *inevitable* outcome of the growth process itself. In particular, in the benchmark model of Galor & Weil (2000) technological progress occurs even for zero education investments and small populations, with the result that eventually Malthusian stagnation vanishes endogenously. Accordingly, the main mechanism of UGT is the scale effect of population, so that the greater the population size (or the population density), the greater the technological change in the Malthusian Epoch and consequently the sooner the economy crosses the threshold level of technological rate of change above which investment in human capital is beneficial to fostering the transition toward the Modern Growth Era. The problem is that historical data regarding the relative sizes and densities of the different regions of the world before any country escapes the Malthusian trap contradicts the plausibility of this most important mechanism of UGT. Indeed, Acemoglu (2009, p.114) notes that the European population has consistently been less than that

---

[1] Before the work of Ashraf, Galor and colleagues was known, Benjamin et al. (2012) wrote an interesting article on the promises and pitfalls of this emerging field of research baptized *Genoeconomics*. Wade (2014) published a book to expose this theory to a wider audience and came in for quite harsh criticism from Ashraf & Galor (2016).

of Asia for the past 2000 years, thus according to the most crucial mechanism of UGT, East Asia should have taken-off before Western Europe. Since the contrary happened it is unlikely that simple scale effects from the population are responsible for the early economic take-off in Western Europe while Asia stagnated. One might argue that in Galor & Moav (2002), the scale effect is no longer mandatory for a take-off to occur since in this model it is the inevitable selection of quality oriented households and their gradual predominance within the whole population that spur the transition from stagnation to growth. Yet, if the population size does not matter in this model it is because of the assumption that the costs (not related to education) of rearing a child do not depend on the population size. Nevertheless, population density is known to have an impact on childrearing costs that are unrelated to education. Specifically, according to De la Croix & Gosseries (2012) citing evidence from Goodsell (1937) and Thompson (1938), when households have small dwellings, child production is more costly and households have fewer children.

### Technological losses and stagnation trap

Moreover, in both Galor & Weil (2000) and Galor & Moav (2002), there can be differential timing and magnitudes of take-off across countries as the result of deep-rooted factors (of a biogeographical, cultural, institutional, or accidental nature), but a country cannot be locked in a stagnation trap, i.e. be maintained in the Malthusian Epoch. Dao & Davila (2013) argue that this is mostly because technological losses are not possible in these models, whereas in reality technology must not only be acquired but maintained too and history provides many examples of technology losses for various reasons. In particular, historians provide evidence that some societies show no sign of escaping stagnation on their own due to losses of technology and culture as a result of geographical isolation on a small and remote island or after a major political reversal that gives rise to isolationist attitudes even in continental economies. In the first case (geographical isolation on a small island), the most famous example, but by no means the only documented one, took place in Tasmania. At the time of arrival of Europeans in the seventeenth century, Tasmanians had the simplest material culture and technology of any people in the world. They were hunter-gatherers but they lacked many technologies and artifacts that were widespread in Aboriginal communities of the Australian mainland, like fishing nets and awls, bows and arrows, and other bone tools. It has been demonstrated that these technologies were present in Tasmania when it was still a part of the Australian mainland, and were subsequently lost when Tasmanians became isolated 10,000 years ago when the sea rose to form the Bass Strait. Diamond (1997, pp.312–13) argues that a small population fluctuating between 3,000 and 7,000 individuals was able to survive for 10,000 years, but was not large enough to prevent significant losses of technology and culture, or to invent new technology, leaving it with a uniquely simplified material culture. Concerning the second source of technology losses (fads that see economically useful technologies becoming devalued), two famous examples are thought to have had long-lasting repercussions. The first one is the loss of gun production technologies in isolated Japan under the Shogunate, when the Samurai class worked against the acceptance of firearms because of a cultural preference for swords as class symbols as well as works of art (Diamond 1997, pp. 257–258; Landes 1998, p.358). The second famous instance of technology losses due to isolationist political attitudes is the abandonment of oceangoing navigation techniques by the Chinese Ming dynasty during the fifteenth century

(Morris 2010, pp.413–417). Dao & Davila (2013) extend the Galor-Weil model to take into account this possibility of growth take-off hampered by technology losses as a result of (geographical or political) isolation. In their model they make abstraction of the causes of technology losses and, for the sake of simplicity, assume a (not necessarily high) rate of recurrent technology losses $\lambda(\omega)$ that depends positively on the degree $\omega$ of isolation of the society.[1] The main conclusion of this model is that in the early stages of development, i.e. with a small population and a low technological level giving households no incentive to educate their children, if the geographical factors, i.e. amount of natural resources, their suitability to the economic process, and the degree of isolation of the environment, do not allow for a sufficiently large population, there will be no technological growth in the long run, as well as no investment in education to enhance technological progress. As a consequence, the economy will be locked in stagnation.

### Lag of demographic transition and the quantity-quality trade-off

The third most important criticism of UGT concerns the central mechanism of quality-quantity tradeoff and the connection between the transition to sustained growth and the demographic transition. For Clark & Cummins (2016) the “crucial underlying assumption that the more children a given set of parents have, the less productive the children will be rests on the flimsiest empirical evidence”. Analyzing English data covering the period 1780–1880, these authors find that family size had no effect on education, occupation, longevity, or even on wealth, though in this case it is wealth at death relative to wealth inherited. Acemoglu (2009, p.736) points out that “there is relatively little direct evidence that this trade-off is important in general or that it leads to the demographic transition”. And indeed, if some studies such as Cáceres-Delphiano (2006) for the US and Li et al. (2008) for China find the expected negative family size–child quality relationship, other empirical studies find no evidence of a quality-quantity trade-off, such as Angrist et al. (2010) for Israel, Black et al. (2005) for Norway, and Clark & Cummins (2016) for England. These last authors conclude that “Modern growth consequently cannot be explained by a switch to smaller family sizes accompanied by more investment in child quality. Modern growth in England began 100 years before there were significant reductions in average family sizes”.

This one hundred year delay between the ignition of the Modern Growth Era and the onset of the demographic transition in England suggests that alternative mechanisms should be called on to better relate the transition from stagnation to growth to the apparently subsequent (and not simultaneous) demographic transition. Twenty years before his theory relating the quantity-quality trade-off of households and the rise in demand for human capital, Becker (1960) advanced the much simpler argument that the decline in fertility was a by-product of the

---

[1] One may also want to make this rate dependent on the society’s population size, education level, technological level, etc. Indeed, a larger and better educated population may be better at maintaining technological knowledge due to dissemination scale and interaction among people. For example, Aiyar et al. (2008) focus on a different phenomenon of technology regress based on external shocks reducing the population in societies in which the transmission of technology is embodied in human capital instead of recorded. They argue that, when the population shrinks, aggregate demand falls, leading some technologies to become unprofitable at the margin. As a consequence, those out-of-use technologies are not transmitted to the next generation, and hence lost until rediscovered by chance. Nevertheless, in the Galor-Weil setting, these effects can be captured in the technological progress function $g_{t+1} = g(e_t; L)$. And indeed, a high technological level itself may spare a society from losing technologies in two ways: (i) through better storage devices in which to save technologies, and (ii) better communications and transportation to offset isolation. But anyway, Dao & Davila (2013) focus on societies in the very early stages of development without widespread literacy and modern communications.

rise in income and the associated rise in the opportunity cost of raising children. This theory hinges on the supposition that individuals' preferences reflect an innate bias against child quantity beyond a certain level of income. Another popular alternative mechanism to explain the demographic transition is to assume that early economic growth gradually decreases mortality rates and consequently improves life expectancy, in such a context the period over which individuals receive their returns on human capital investment is prolonged, which increases the rate of return on human capital and hence launches the economy on a virtuous circle of human capital investment, declining fertility, and sustained economic growth (Boucekkine et al. 2003; Cervellati & Sunde 2005; Soares 2005). Hazan & Zoabi (2006) support the alternative idea that health improvements (but not longevity gains) associated with increasing income are more likely to generate a quantity-quality trade-off and the consequent demographic transition and investment in human capital. Religious movements and the Age of Enlightenment also seem to have contributed to the demographic transition through their effects on preferences for education (Becker et al. 2010). Other social scientists proposed that social norms that have reduced the gender income gap (Lagerlöf 2003) and the need for child labor (Doepke 2004), or more simply the establishment of capital markets that have decreased the need for parents to be taken care of by their children in their old age (Neher 1971; Zhang & Nishimura 1993), have been essential to triggering the demographic transition and the rise in human capital. Empirical evidence can be found either to support or denigrate each of these assumptions, so that as of yet there is no general consensus on the causes of the demographic transition or the role of the quality-quantity trade-off in determining population dynamics.

In order to have a more complete description of the economic growth process, this second chapter has concentrated on the study of mainstream economic theories that focus on the state of sustained growth attained by industrialized countries. Despite a considerable literature, the proximate causes of economic growth of these theoretical models are always the same. They consist in the accumulation of physical and human capital, and the improvement of the efficiency of the economy to grow output from these inputs. This last variable is called technological change by mainstream economists and corresponds to a catch-all aggregation of many different features of the economic system that, for most of them, should in fact be partially regarded as consequences, or as facilitating factors, of economic growth. The proximate factors on which conventional economic theory focuses cannot be fundamental causes of growth, for the simple reason that "at some level (and exaggerating somewhat) to say that a country is poor because it has insufficient physical capital, human capital, and inefficient technology is like saying that a person is poor because he or she does not have money" (Acemoglu 2009, p.106).

If the proximate causes of growth (physical and human capital accumulation, technological change) remain the same in all the different approaches presented in this chapter, the way they are interlinked in Unified Growth Theory is more ambitious. Indeed, compared to the more conventional settings of neoclassical economics that focus on the Modern Growth Era, UGT is a far more powerful framework in which to investigate the transition from stagnation to sustained growth. UGT has undoubtedly brought a new perspective to the process of economic growth, but its core mechanisms are clearly questionable. UGT will be a more promising avenue to a fuller understanding of long-term economic growth once mechanisms developed on a stronger empirical basis are incorporated into its analytical framework.

As the next chapter will show, the most important aspect that must be incorporated into UGT is the physical reality of the economic system. As will be exposed, only useful exergy consumption can constitute an ultimate cause of economic growth.

# CHAPTER 3

# ENERGY, ENTROPY, AND ECONOMIC GROWTH

> *"Just as the constant increase of entropy is the basic law of the universe, so it is the basic law of life to be ever more highly structured and to struggle against entropy."*
>
> Václav Havel

The role of energy is absolutely insignificant in most standard textbooks on economic growth theory. The term "energy" does not feature a single time in Jones (1998), Aghion & Howitt (1998), de La Croix & Michel (2002) or Barro & Sala-i-Martin (2004). In Acemoglu (2009) and Aghion & Howitt (2009) energy is mentioned in relation to just one econometric study by Popp (2002) who investigates innovation in energy sectors.[1] The present chapter first presents the three main arguments of neoclassical economics for ignoring the role of natural resources, and in particular energy, in the economic process. The value theory endorsed by neoclassical economists is a first explanation for the minor role ascribed to energy. But mostly it is shown that two other postulates of mainstream economics, namely the so-called *cost share theorem* and the supposed high degree of substitutability of natural resources with human-made capital, are based on tenuous (if not fallacious) reasoning. Then, essential concepts such as exergy and entropy are presented in order to explain why the first and second fundamental laws of thermodynamics always apply to the economic system and shape its functioning. Taking into account these universal laws and their associated concepts provides a very different theoretical framework within which to study the economy. This approach is called *biophysical economics* or *thermoeconomics*. The implications of this paradigm for the analysis of long-term societal development are investigated in a third part. Useful exergy consumption appears as the primary determinant of human societal development, and systems of energy capture (foraging vs. farming vs. fossil fuel burning) largely orient the values held by people when interacting with each other. Furthermore, a biophysical approach to the economy demonstrates that the expected *dematerialization* of the economy could be an illusion.

---

[1] The less mathematically formalized and more historically oriented book by Weil (2013) does a slightly better job than other textbooks since it mentions energy several times, essentially in the context of the Industrial Revolution.

## 3.1 MISGUIDED OMISSION OF ENERGY FROM MAINSTREAM ECONOMICS

### *3.1.1 Shift in Value Theory and Loss of Biophysical Reality*

From the classical production cost orientation to the neoclassical utilitarian vision

The formulation of a value theory is generally essential to the emergence of a new school of thought in economics. In a sense, it is the primary step on the path to proposing an economic paradigm. For Costanza (2004, p.337), "value systems refer to intrapsychic constellations of norms and precepts that guide human judgment and action. They refer to the normative and moral frameworks people use to assign importance and necessity to their beliefs and actions". Classical economists formulated a production-cost-based theory of value, in which wages, profits, and capital annuities were the three original sources of exchange value. For Smith (1776), the real price of goods is set by the amount of human labor that has to be commanded to produce the good. For him "labor, therefore, never varying in its own value, is alone the ultimate and real standard by which the value of all commodities can at all times and places be estimated and compared". Smith (1776) uses a famous and caricatural example to illustrate his point. If you imagine that killing a beaver requires twice the labor required to kill a deer, logically, one beaver will sell for as much as two deer. As already highlighted, this view (developed in more subtle and different detail by Ricardo and Marx) vanished when the so-called *marginal revolution* was operated. Menger (1871) proposed that there are different categories of wants and desires, such as food, shelter, and clothing. These categories can be ordered according to their subjective importance for people. Furthermore, within each category, one can define "an ordered sequence of desires for successive increments of each good" (Costanza 2004, p.339). Then, Menger (1871), Jevons (1971), and Walras (1974) theorized the economic principle of diminishing marginal utility when they advanced that, within a given good or service category, the intensity of desire for one additional unit declines as the cumulative number of consumed units increases. The transfer that occurred from the production-oriented value theory of the classical economists to the consumption-oriented value theory through preference of the Marginalists is clear in the words of Costanza (2004, p.340): "while the classical theorists sought a standard physical commodity unit for measuring exchange value, neoclassical theorists substituted utility for such a commodity. Because value was assumed to be determined solely by consumption utility on the margin, and consumers were assumed to allocate money optimally across uses (in possession of perfect information, no externalities, fixed preferences, and no interpersonal effects), the marginal utility of money was the same for an individual in all its uses. Money thus became the standard unit of measure".

The consequent loss of biophysical ground

This shift explains why standard economics is now focused on the general optimization problem of labor vs. leisure and consumption vs. saving while respecting time and wealth constraints. People are assumed to be rational and make their decisions in order to optimize their utility given a constraint on income and/or time. As a consequence, time or money can be used as standard measures of use value, so that the problem that needs to be solved looks like: How much time or money will a person willingly sacrifice to obtain commodity $X$? In this context, the world is largely deterministic and moves from one equilibrium to another in a relatively stable and predictable fashion. The determination of this equilibrium is a result of the conflicting forces operating at the same time such as supply, demand, unlimited wants and limited means. As can be seen, even though value can be more generally extended to depict the contribution to a goal or an objective, the neoclassical approach makes value a representation of want, satisfaction, and pleasure, i.e. utility. As a result, it also makes this mental representation intrinsically anthropocentric (Costanza 2004, p.340). In this view, the economic system is observable for itself and is not subject to the physical laws of the broader environment because a myopic human centered view is adopted. As a consequence, neoclassical theory treats macroeconomic phenomena as the sum of the individual decisions based on tastes and preferences and the objective for neoclassical economists is to ensure that markets operate efficiently. In such a framework there is no place for biophysical constraints since the behavior of the economic system is independent of the physical state of its surrounding environment. Hence, the value theory of mainstream economics is the first reason why energy, and natural resources more generally, are of no immediate concern to neoclassical economics.

#### Value system and the endless debate between Doomers and Cornucopians

The importance of value beliefs is clear in *The Limits to Growth* controversy. In 1972, a research team from MIT released a book called *The Limits to Growth* (Meadows et al. 1972) whose goal was to explore the complex feedback nature of the global economy-environment system more than it was to predict anything in particular. The thing is, most of the scenarios run with the simulation model supporting the study (called World3) displayed an overshoot and collapse of the global economy. The book immediately came in for much criticism from economists (Kaysen 1972; Solow 1973). The most complete review came from a research group at the University of Sussex in the form of a book called *Models of Doom: A Critique of the Limits to Growth* (Cole et al. 1973). The purpose of this section is not to analyze the different arguments of this controversy (a comprehensive review is provided by Bardi 2011), but the interesting point here is to note that Meadows et al. had the opportunity to respond to the different criticisms of Cole et al.. At the end of their discussion, the authors reached the conclusion that it was mostly their core beliefs that differed and prevented them from further responding to each other on a rational and fruitful basis. Broadly speaking, people will always be divided between pessimistic *Doomers*, who think that human systems are by essence subjected to the natural broader environment, and optimistic *Cornucopians*, who think that mankind’s unique ingenuity enables us to repeatedly overcome environmental constraints through technological progress.

### *3.1.2 Aggregate Production Function and Cost Share Theorem*

#### Criticisms of the very notion of aggregate production function

As already emphasized, the basic structure of the neoclassical economic growth theory depends on the notion of a differentiable and homogenous macroeconomic production function whose arguments are stocks of production factors. The very concept of aggregate production function has been much criticized from its first uses to the present day (Fisher 1969; Pressman

2005; Felipe & McCombie 2005). Lindenberger & Kümmel (2011) resume the three principal objections leveled at the concept of aggregate production function. First, the problem of aggregating heterogeneous output goods and services into one monetary quantity, measured in national accounts by deflated GDP. Second, the problem of aggregating the heterogeneous components of the capital stock (machines, buildings, etc.) into one monetary quantity called capital, measured in national accounts in deflated currency. Third, the unclear relationship between the micro theory of production in individual firms, for which the concept of the (micro) production function is not questioned, and the macro theory of production.

### The cost share theorem and its refutation

But if it is accepted that there is a macroeconomic production function, it can be shown that the total economic product is used up if each factor is rewarded according to its marginal product, which is automatically the case with a homogenous production function of order unity (i.e. linearly homogenous) according to Euler's theorem (Blaug 1985). In neoclassical economics *the output elasticities of production factors are equal to their cost share* (i.e. their income allocation) in total GDP since this yields an extremum after Euler-Lagrange optimization. And since national income is allocated almost exclusively to capital and labor it seems logical to postulate that only these factors are significant for production. To be more specific, in industrialized countries such as the USA, labor wages capture around 70% of the total national income, while incomes from capital (i.e. capital rents) represent 30% of this same national income. Hence, in neoclassical production functions such as the Cobb-Douglas function, the output elasticities (i.e. marginal productivities) of capital and labor are respectively 0.3 and 0.7. On the contrary, the cost of energy is on average around 5–6% of the total factor cost (i.e. 5% to 6% of the total national income) in advanced industrial market economies with modern growth regimes. Thus, economists tend either to neglect energy as a factor of production or they argue that if energy is to be accounted for as a production factor, its marginal productivity should logically equal its small cost share of 0.05 (Denison 1979). The cost share theorem is the second reason why neoclassical growth theory puts a particular emphasis on (quality adjusted) capital and labor, and pays far less attention to energy (and other natural resources such as land and materials).

As presented in Appendix C (from the initial work of Kümmel 1982; and later Kümmel et al. 1985; 2002; 2010; and Lindenberger & Kümmel 2011), this cost share theorem is in fact completely inaccurate when two proper *hard*[1] technological constraints are taken into account: (i) the degree of capacity utilization of aggregate capital cannot exceed one, and (ii) the degree of automation cannot exceed the technologically possible degree of automation of one. Appendix C shows that when these technological constraints are taken into account, shadow prices that add to factor costs mean that the cost share theorem no longer holds. According to these authors, using CES or Cobb-Douglas functions to represent economic growth would not be so inaccurate if there were no such technological constraints. But the thing is that they do exist and must be accounted for in any economic model. An empirical test of the cost share theorem was provided by the two oil crises of the 1970s. During the first energy crisis of 1973

---

[1] For Ayres et al. (2013) there are also some *soft* constraints that correspond to social, financial, organizational, or legal restrictions, which limit substitution possibilities between inputs over time as measured empirically (see Section 3.1.3).

when the quantity of energy consumed in the US economy fell by as much as 7%, mainstream growth theory predicted that macroeconomic output should have decreased by no more than a fraction of a percentage point (0.35% to be exact) whereas in reality output in industrialized countries fell by as much as 5%. This empirical test clearly showed, in 1973 and a few years later during the second energy crisis of 1979, that the output elasticity of energy (of 0.05) is clearly underestimated and that consequently the elasticities of capital and/or labor are clearly overestimated. Giraud & Kahraman (2014) confirm this intuition empirically with several important results. First of all, they assert that *output elasticities* cannot readily be measured and that econometric studies should focus instead on *dependency ratios*. Indeed, the output elasticity of primary energy $E$ is the percentage change in output $Y$ (GDP) per one percent change in input factor $E$, *all other production factors* $(K, L, A)$ *remaining constant*; whereas the dependency ratio of GDP with respect to primary energy $E$ is the percentage change in output $Y$ per one percent change in input factor $E$, with *other production factors* $(K, L, A)$ *not being held constant since they depend on energy consumption.* Hence, "only when all the other input factors are independent from energy will the dependency ratio and output elasticity coincide". Following this logic, Giraud & Kahraman (2014) show econometrically that for 33 countries between 1970 and 2011, the dependency ratio of GDP with respect to primary energy lies on average between 0.6 and 0.7, more than ten times the output elasticity of energy of 0.05 traditionally assumed in neoclassical models. Despite the obvious inability of the usual production functions to represent reality, economists have not changed their approach and continue to support the cost share theorem.

### *3.1.3 Substitutability of Natural and Human-Made Input Factors*

Elasticity of substitution between input factors

Let us recall that the elasticity of substitution, noted $\sigma$, between human-made physical capital and an abstract natural resource input (aggregating materials such as metals, energy, ecosystem services, etc.) expresses by how much one of the two inputs must be increased to maintain the same level of output when the use of the other input is reduced (Stern, 2004). The higher $\sigma$, the easier it is to substitute inputs. For instance, with *perfect substitutability* between inputs ($\sigma = 1$), as the ratio of the two inputs is changed by a given percentage (holding output constant), the ratio of their marginal products changes by the same percentage and in the opposite direction (Stern, 2004). Figure 3.1 presents different output isoquants (i.e. the output is constant along the curve) with three particular cases of elasticities of substitution between inputs:

- Perfect substitutability with $\sigma = 1$: as resource use decreases toward zero, production is maintained by increasing capital use towards infinity.
- Infinite substitutability with $\sigma = \infty$: Resources and capital are equivalent; producers see no difference between the two inputs and thus use the cheapest one only.
- Zero substitutability with $\sigma = 0$: inputs are complementary and must be used in a fixed ratio.

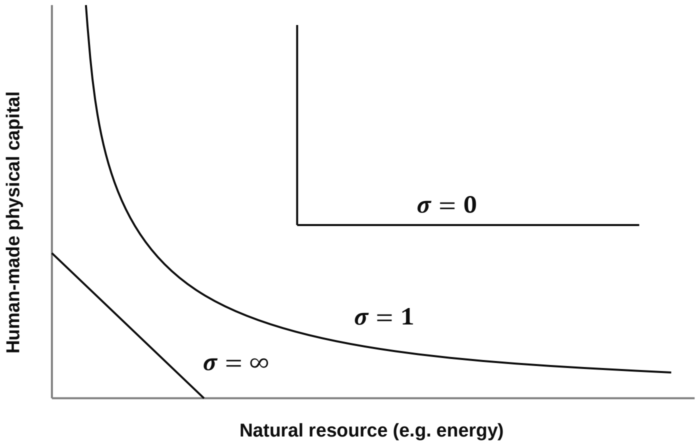


**Figure 3.1 Substitutability in aggregated production function.**
**Source: reproduced from Stern (2004).**

In neoclassical economics, if $\sigma > 1$, natural resources are said to be *non-essential*. With $0 \leq \sigma \leq 1$ natural resources are said to be *essential*. Essential means that considering positive manufactured capital input, output production is only zero when the resource input is zero, and strictly positive otherwise (Stern, 2004). An example of a function respecting the essentiality condition is the Cobb-Douglas production function where $\sigma = 1$. Economists argue that using such a production function accounts for the fact that some amount of energy and material is always required in production processes and thus such a description of production respects the mass balance and thermodynamic constraints presented in Section 3.2. In fact, in such a production function energy or natural resources can tend to infinitesimal quantities whereas thermodynamic constraints stipulate that limits are surely far higher (i.e. non-infinitesimal) as shown in Section 3.3.3.

The non-essentiality of the abstract natural resource input in neoclassical models

After the publication of *The Limits to Growth*, neoclassical economists started to incorporate an *abstract natural resource input*[1] into their production function (Solow 1974; Stiglitz 1974; Dasgupta & Heal 1974). Most of the time such efforts were made in the perspective of sustainability discussions with optimal growth models that attempted either to maximize the sum of discounted social welfare over an infinite time horizon or to achieve non-declining social welfare (Stern 2004). These models focused especially on institutional arrangements such as market structures (competition vs. central planning), the system of property rights (private vs. common property), and the system of values with respect to the welfare of current and future generations (intergenerational vs. intragenerational equity). But in such models, technical conditions defining the mix of renewable and non-renewable resources, the initial state of capital and natural resources, and the ease of substitution among inputs that should be met to permit sustainability are in fact *a priori* assumed to be feasible and thus not questioned. As a consequence, institutional conditions are the only questions that are truly investigated by neoclassical economists in such models (Stern, 2004).

---

[1] The natural resource input is loosely defined in these models. It does not specifically represent energy but rather the general *services* that the environment provides to mankind free-of-charge. It is commonly modeled as a renewable stock, i.e. a depletable stock that possesses a natural capacity to regenerate.

More precisely, in neoclassical growth models such as the ones described in Solow (1974), Stiglitz (1974), and Dasgupta & Heal (1974), $\sigma$ must be equal to or greater than 1 for sustainability to be at least technically feasible. In that case, society *has just* to invest sufficiently rapidly in physical capital to replace the depleted natural resources in order to achieve sustainability. In reality, the empirical review of van der Werf (2008) shows that the elasticity of substitution between energy and human-made (physical or human) capital is closer to 0.5, implying that these production factors are not readily substitutable. The presupposed high substitutability between natural resources (and in particular energy) and physical capital is the third reason for neglecting energy in the growth process.

## 3.2 THEORETICAL FRAMEWORK OF BIOPHYSICAL/THERMO ECONOMICS

### *3.2.1 Repeated Neglect of Available Wisdom and Definitions of Energy and Exergy*

Repeated neglect of available wisdom

The fact that the role of energy is still improperly taken into account in economic theory nowadays is even more troubling when one thinks that critical opinions on this issue have been raised for more than one hundred and thirty years. Indeed, after enumerating all the different sources of energy (air and water movement such as wind and waves, biomass and internal Earth heat) from which humans can extract primary energy, Podolinsky (1880) asserts that “all these examples clearly demonstrate that the radiated energy from the sun is almost the unique source of all beneficial forces of humanity that are on the surface of Earth”. Around the same time, Spencer (1890) claimed that “when we contemplate a society as an organism, and observe the direction of its growth, we find this direction to be that in which the average of opposing forces is the least. Its units have energies to be expended in self-maintenance and reproduction”. For Ostwald (1911, p.870), “the progress of science is characterized by the fact that more and more energy is utilized for human purposes, and that the transformation of the raw energies […] is attended by ever-increased efficiency”. Soddy (1926) dedicated a large part of his professional career to elaborating a critique of the standard economic theory and claimed that “if we have available energy, we may maintain life and produce every material requisite necessary. That is why the flow of energy should be the primary concern of economics”. Mostly, for Tryon (1927) “anything as important in industrial life as power deserves more attention than it has yet received from economists…A theory of production that will really explain how wealth is produced must analyze the contribution of the element energy”. Other scholars have made important contributions to try to rehabilitate the role of energy in economic theory and their work will be discussed later on. The previous examples were given to show how most of economic growth theories have been continuously hermetic to these successive criticisms and failed to question its artificial conception of the economic system.

Definition of energy

We must now do something that has been avoided so far but that is indispensable if we are to move forward: we must give a clear physical definition of energy. This apparently simple task is in fact quite complicated because it implies reference to other physical concepts. In the

International System of Units, energy is measured in joules and 1 joule (J) is defined as the quantity of *mechanical work*, noted $W$, transferred to an object by moving it a distance of 1 meter (m) against a force of 1 newton (N).[1] In this sense, energy is something that is *transferred* as work. Another familiar mode of energy transfer is *heat*, noted $Q$, and in that context 1 joule is equal to 0.2389 calorie (cal) where 1 calorie represents the energy needed to raise the temperature of one gram of water by one degree Celsius at a pressure of one atmosphere.[2] From a molecular viewpoint, the motion of an object affected by mechanical work corresponds to all its atoms moving in the same direction. Hence, *work is the transfer of energy that makes use of the uniform motion of atoms in the surroundings*. On the contrary, two juxtaposed blocks of the same material (say iron) at different temperatures will exchange heat from the hotter to the colder block. Initially, the atoms of the hotter block oscillate around their average positions more vigorously than those in the cooler block, but if the two blocks are brought into contact long enough, their atoms will eventually oscillate to the same extent around their average positions. Hence, *heat is the transfer of energy that makes use of the random motion of atoms in the surroundings* (Atkins 2010, pp.24–25). If mechanical work and heat are the most familiar modes of *energy transfer*, energy is mostly a property of systems defined by precise components and boundaries. *System energies* include *kinetic energy*, which is the energy of motion; *potential energy*, which is the energy of a mass in a gravitational field (with *coulomb energy* as the potential energy of a charge in an electric field); *electric and magnetic energies*, which are related by Maxwell's equation to coulomb energy; *photon energy*, which is the energy of an electromagnetic wave such as light; and *chemical energy*, which is the internal energy of a system of many interacting particles. Depending on context, a given system can be defined by one or several forms of the previously described energy $E_i$, so that its internal energy $U$, which is the exact measure of the *quantity* of energy that this system possesses, is the sum of all of them:

$$U = \sum_i E_i. \tag{3.1}$$

Definition of exergy

In the early stages of the Industrial Revolution, people noticed that as well as varying in quantity, industrial processes also varied in *quality*. More precisely, the fraction of energy that can be converted into mechanical work is not the same from one process to another. This led scientists to differentiate between two components of a given energy quantity: a useful part, called *exergy* and noted $E_X$, which can be converted into any type of physical work,[3] and a useless part, called *anergy* and noted $E_A$, which is transferred as low heat to the surrounding environment during system changes (industrial processes). Hence, the quality of a given quantity of energy changes according to its relative fractions of exergy and anergy, but the sum of the two make energy a constant:

---

[1] Hence, 1 J = 1 N.m, and 1 N is the force needed to accelerate one kilogram of mass at the rate of one meter per second squared in the direction of the applied force, i.e. 1 N = 1 kg.m$^2$.s$^{-2}$.

[2] The standard atmosphere (atm) is a unit of pressure defined as 101325 Pascal (Pa), or 1.01325 bar.

[3] Formally, exergy is the maximum amount of work that can theoretically be recovered from a system as it approaches equilibrium with its surroundings reversibly, that is, infinitely slowly (Ayres & Warr 2009).

$$Energy = Exergy + Anergy = Constant. \tag{3.2}$$

Throughout industrial processes, energy is always conserved but exergy is gradually destroyed and converted into anergy because each step in an industrial process occurs with *irreversibilities* at the micro scale, which are visible as frictions and heat losses at the macro scale. Generally, these released heat outflows have higher temperatures than the wider environment so they still contain some exergy. As they gradually mix with the surrounding environment, the temperature of these heat losses ultimately equals the temperature of the environment and accordingly their exergy content (i.e. their capacity to do work) decreases, so that ultimately they end as complete anergy. Thus, in the conversion processes, energy is conserved in quantity but its quality declines as it gradually loses all its ability to perform work: anergy increases at the expense of exergy (Kümmel 2011, p.114). For example, think about coal extracted from a given deposit and burned in a power plant to produce electricity used to light a bulb. At the beginning of this industrial process, the chemical energy stored in coal is about 98% exergy and 2% anergy. When burned in the power plant, this given quantity of coal will produce about 30% of electricity, which is 100% exergy, and 70% of waste heat released into the environment, which contains some exergy but eventually completely dissipates as anergy. Finally, the electricity sent to the light bulb is converted into light measurable in terms of exergy and represents the true useful exergy service provided to people, but all this light is ultimately dissipated as heat in the environment, whether at the surface of the bulb or through the radiant electromagnetic waves of light. Hence, the energy quantity extracted as coal at the beginning of the process ends in the exact same quantity as heat at the end of this industrial process. However, at the beginning this energy was almost 100% exergy but at the end it is necessarily 100% anergy because exergy is gradually destroyed.[1] When we speak about *energy consumption* we are in fact referring to *exergy consumption* because by definition energy cannot be consumed.

### *3.2.2 Entropy and Fundamental Laws of Thermodynamics*

Definition of entropy

The gradual depreciation of the quality of energy, i.e. the gradual destruction of exergy, is part of an irresistible tendency of all natural and technical systems to spread out their components as evenly as possible in space and over the states of motion (Kümmel 2011, p.114). In other words, systems move naturally toward their most *disordered state* in the absence of work available to maintain order. *Entropy, noted $S$, is the physical measure of disorder and all energy conversion processes produce entropy*. For instance, when coal is burned in a furnace, the constitutive atoms of the resulting gases, particles, and ashes are organized in a less ordered state (high entropy) compared to their initial highly ordered state (low entropy) in solid coal. In this example, but also in general, entropy production is coupled to emissions of heat and particles. From a molecular point of view, a statistical mechanics approach is needed to

---

[1] This fact demonstrated above in the case of coal burned to produce electricity used as light is true for all other economic processes. Think about crude oil extracted to produce refined gasoline used as car motion. In such a process, energy is conserved (from chemical to mechanical and then heat energy) but exergy is gradually destroyed in the creation of anergy.

understand the concept of entropy as a measure of the number of ways in which a system may be arranged. For a given *macrostate* characterized by plainly observable average quantities of macroscopic variables such as temperature, pressure and volume, the entropy measures the degree to which the probability of the system is spread out over different possible *microstates*. In contrast to the macrostate, a microstate specifies all the molecular details about the system including the position and velocity of every molecule. Hence, the higher the entropy, the higher the number of possible microscopic configurations of the individual atoms and molecules of the system (microstates) which could give rise to the observed macrostate of the system. Specifically, Planck (1903) draws on Boltzmann (1902) to assert that entropy is a logarithmic measure of the number of states with significant probability of being occupied, hence:

$$S = -k_B \sum_i p_i \ln p_i, \tag{3.3}$$

where $k_B$ is the Boltzmann constant, equal to 1.38065×10−23 J.K$^{-1}$. The summation is over all the possible microstates of the system, and $p_i$ is the probability that the system is in the $i$-th microstate. In what has been called *the fundamental postulate in statistical mechanics*, the occupation of any microstate is assumed to be equally probable, i.e. $p_i = 1/\Omega$, where $\Omega$ is the number of microstates. Then, the previous equation defining entropy $S$ (the physical measure of disorder of the system) reduces to

$$S = -k_B \ln \Omega. \tag{3.4}$$

In classical (and not statistical) thermodynamics, entropy is a state function, meaning that it does not depend on the path by which the system reached its present state. According to Clausius (1867), the amount of entropy change $\Delta S$ of a given system is the energy reversibly transferred as heat $\Delta Q_{rev}$ divided by the (absolute[1]) temperature $T$ at which the transfer took place:

$$\Delta S = \frac{\Delta Q_{rev}}{T}. \tag{3.5}$$

Atkins (2010, p.48) provides a very useful analogy to explain the concept of entropy and to see the importance of the temperature $T$ at which the heat transfer $\Delta Q_{rev}$ takes place. Imagine a quiet library as a metaphor for a system at low temperature with little thermal motion. In such a context, if someone with a very bad cold sneezes suddenly it will be highly disruptive for the other people in the library: there is a sudden large increase in disorder, i.e. a large increase in entropy. On the other hand, a busy street is a metaphor for a system at high temperature with a lot of thermal motion. Now the exact same sneeze will introduce relatively little additional disorder in the busy street: there is only a small increase in entropy. In each case the additional disorder, or increase in entropy, is proportional to the magnitude of the sneeze (the quantity of

---

[1] The absolute or thermodynamic temperature uses the Kelvin (K) scale of measurement and selects the triple point of water at 273.16 K (0.01°C) as the fundamental fixing point. Like the Celsius scale (but not the Fahrenheit scale), the Kelvin scale is a centigrade scale so that converting between the Kelvin and Celsius scales is simple: T= 0 K = -273°C; T = 273 K = 0°C.

energy transferred as heat $\Delta Q_{rev}$) and inversely proportional to the initial agitation of the library/street (the temperature $T$ of the system).

Entropy increase and exergy degradation

The inherent entropy production, associated with combustion or friction processes which create disorder from order, reduces exergy and enhances anergy. Formally, consider a combustion process that consists of a many-particles system of internal energy $U$, entropy $S$, volume $V$, and pressure $p$, which is out of equilibrium with its environment at temperature $T_0$ and pressure $p_0$. The system can exchange heat, work, and matter with its environment through a current of stationary speed $v$, mass $m$, and height $z$, which enters it at one place and leaves it at another with the kinetic energy $mv^2/2$, and the potential energy $mgz$ where $g$ is the standard acceleration due to gravity.[1] If the system consists of $N$ different sorts of $i$ particles, with $n_i$ being the number of particles of component $i$, and if $\mu_{i0}$ and $\mu_{id}$ are the chemical potentials of component $i$ in thermal and mechanical equilibrium before and after diffusion respectively,[2] then the exergy content of the system is:

$$E_X = U - U_0 + pV - p_0V_0 - T_0(S - S_0) + \sum_{i=1}^{N} n_i(\mu_{i0} - \mu_{id}) + mv^2/2 + mgz. \qquad (3.6)$$

The first and second laws of thermodynamics and perpetual motion machines

With all these concepts in mind, the fundamental laws of thermodynamics and their universal implications can be more easily understood. I formulate here my own version of the first and second fundamental laws of thermodynamics[3] based on Atkins (2010) and Kümmel (2011).

The first law of thermodynamics states that:

> *The total energy of an isolated system is constant, energy can be transformed from one form to another but cannot be created or destroyed. As a corollary, it is impossible to construct a perpetual motion machine of the first kind; that is, a machine that performs work without any input of energy.*

---

[1] The standard acceleration due to gravity $g$ is 9.80665 m.s$^{-2}$.

[2] The concentration of the combustion products in the combustion chamber is higher than that in the environment so in principle work can be obtained from their diffusion into the environment. To compute this work which contributes to exergy, the implicit assumption here is to pretend that immediately after combustion the system components are mixed and already in thermal and mechanical equilibrium with the environment, but that they left the combustion chamber in a second step.

[3] There is a total of four laws of thermodynamics but only the first and second are useful in understanding the economic process. The Zeroth law of thermodynamics states that if two systems are in thermal equilibrium with a third system, they are in thermal equilibrium with each other (this law helps to define the notion of temperature). The Third law of thermodynamics states that the entropy of a system approaches a constant value as the temperature approaches absolute zero, and with the exception of non-crystalline solids (glasses) the entropy of a system at absolute zero is typically close to zero.

The second law of thermodynamics[1] states that:

> *The total entropy of an isolated system increases over time and exergy is necessarily degraded in spontaneous processes due to irreversibilities which produce entropy. As a corollary, it is impossible to construct a perpetual motion machine of the second kind; that is, a machine that does nothing other than cooling down a heat reservoir and performing work.*

Figure 3.2 helps us to understand the relevance of the first and second laws of thermodynamics. As shown in Figure 3.2a, a perpetual motion machine of the first kind ($M_{ppl\,1}$) delivers work without the input of energy. In such a system, energy in the form of work is created *ex nihilo,* which is completely impossible. It is impossible to see a car moving without a given fuel input (be it gasoline, electricity, etc.), just as it is impossible to see electricity coming out of a power plant without seeing the prior entry of coal, oil, or nuclear fissile material into that same power plant. Such a perpetual motion machine of the first kind will never be observed as it contradicts the first law of thermodynamics. As depicted in Figure 3.2b, a simple way to correct the failure of the perpetual motion machine of the first kind appears to be to add a reservoir[2] of temperature $T$ from which the machine could extract and entirely convert some heat $Q$ into work $W$. In such a machine there would be absolutely no friction, no losses, so that $W = Q$, and the reservoir will inevitably cool down as time goes toward infinity.[3] No one has ever seen a machine without friction or losses. Moreover, such a perpetual motion machine of the second kind ($M_{ppl\,2}$) implies an overall decrease in the total entropy of the system, meaning that the entire system is increasingly ordered by the functioning of the machine. A perpetual motion machine of the second kind cannot exist as it violates the second law of thermodynamics. As shown in Figure 3.2c, the second law and practical experiments indicate that a heat engine that receives an amount of heat $Q$ from a donating reservoir at temperature $T$ must always give off a part $Q_0$ of that heat to a second receiving reservoir at a lower temperature $T_0 < T$. Thus, this real motion machine ($M_{real}$) can only perform the work $W = Q - Q_0$. The heat $Q_0$ dissipated to the natural environment contains no exergy and is therefore useless since its temperature $T_0$ is equal to the environment's temperature. Furthermore, as the upper heat-providing reservoir cools down it becomes increasingly ordered and hence its entropy decreases so that $\Delta S_T < 0$, whereas on the other side the lower heat-receiving reservoir becomes increasingly disordered and hence its entropy increases, so that $\Delta S_0 > 0$. The second law of thermodynamics states that this real motion machine is *spontaneous*, i.e. can *deliver* the work *output* $W$, if the total entropy change is positive, i.e. if the rise in entropy in the lower heat-receiving reservoir is greater than the fall in entropy of the upper heat-providing reservoir, that

---

[1] Historically, Clausius (1867) gave the following statement of the second law: "heat does not pass from a body at low temperature to one at high temperature without an accompanying change elsewhere". The version of Lord Kelvin (1911; William Thomson before being made 1st Baron Kelvin) was: "no cyclic process is possible in which heat is taken from a hot source and converted completely into work". These definitions are equivalent as demonstrated by Atkins (2010).

[2] A reservoir is a many-body system whose number of degrees of freedom far exceeds that of all other systems interacting with it. Therefore, any exchange of heat, work, and matter with other systems leaves it in an undisturbed equilibrium at constant finite temperature, pressure, and chemical composition (Kümmel 2011, p.115).

[3] In that sense it violates one of the properties of a reservoir, which is to see its temperature undisturbed and constant as explained in the footnote just above.

is if $|\Delta S_0| > |\Delta S_T|$. If the sign of this inequality is reversed while the lower reservoir is still at temperature $T_0$ and the upper one at temperature $T > T_0$, it means that heat flows from the lower reservoir to the upper one and that the overall process is *not spontaneous* and necessitates the *input* of work $W$. Such a non-spontaneous process is possible only if it is coupled to a spontaneous process that occurs somewhere else such that the total entropy of both coupled systems is indeed increasing. For example, cooling food in a refrigerator is a non-spontaneous process that is possible only if a spontaneous process, fossil fuel combustion for instance, occurs in a remote power plant somewhere to deliver work in the form of electricity to the refrigerator.

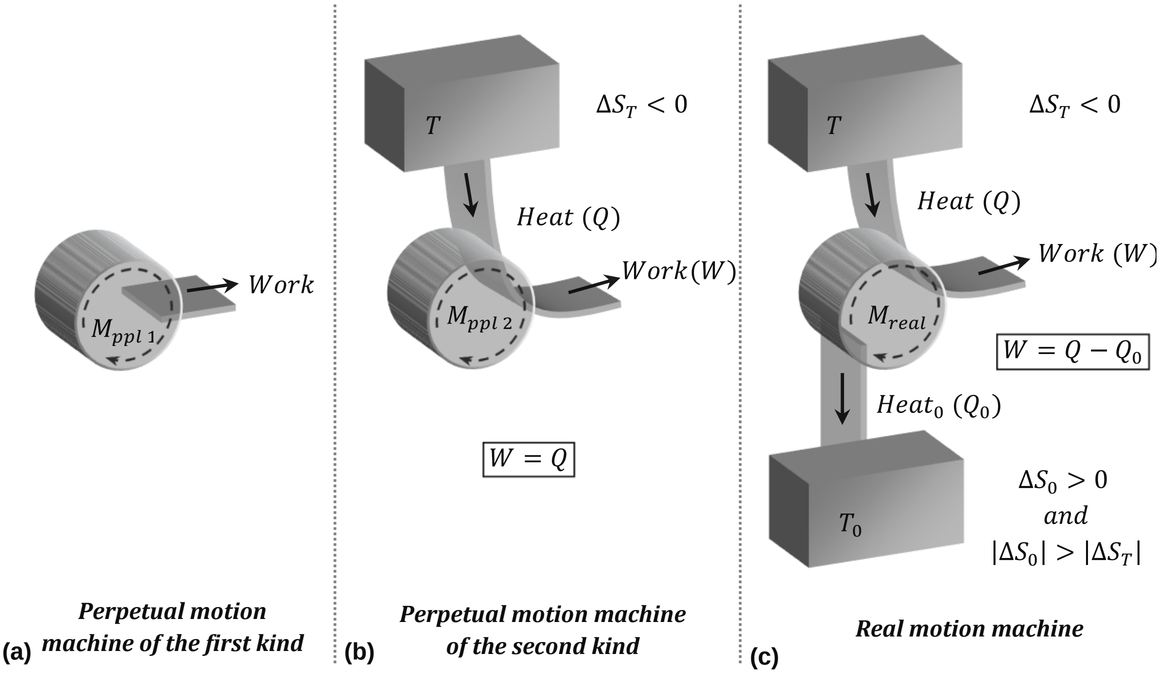


**Figure 3.2 The compelling necessity of the first and second laws of thermodynamics.**
**Differences between (a) perpetual machine of the first kind, (b) perpetual motion machine of the second kind, and (c) real motion machine. Source: adapted from Atkins (2010) and Kümmel (2011).**

It is important to see the complementarity of the two fundamental laws of thermodynamics as described by Atkins (2010, p.51). The first law and the internal energy identify the *feasible* change among all conceivable changes: a process is feasible only if the total energy of the universe (system under study + surrounding environment) remains constant. The second law and entropy identify the *spontaneous* changes among the feasible changes: a feasible process is spontaneous only if the total entropy of the universe increases. On this last point it is crucial to stress again that entropy can decrease locally for a given system, but the price of increasing local order is necessarily enhanced disorder in the broader environment and loss of energy quality during the process of local increasing order (Kümmel 2011, p.114). Moreover the increase in disorder of the broader environment is necessarily higher than the increase in order locally, so that overall the total entropy of the universe[1] is indeed always increasing.

---

[1] Regarding not the universe in the thermodynamic analytical sense (i.e. the system under study plus its broader environment playing the role of a cold heat-receiving reservoir) but the whole universe we live in, Kümmel (2011, p.118) explains that one

The arrow of time

Before applying the fundamental laws of thermodynamics to the economy-environment system, a further word is needed on the general implications of the first and second laws. This discussion concerns the uniformity and the direction of the arrow of time. First, energy is conserved as required by the first law of thermodynamics because time is uniform: time flows steadily, it does not bunch up and run faster then spread out and run slowly. Time is a uniformly structured coordinate. If time were to bunch up and spread out, energy would not be conserved. Thus, the first law of thermodynamics is based on a very deep aspect of our whole universe (Atkins 2010, p.35–36). Second, the flow of time has a one-way direction or *asymmetry*. The *psychological arrow of time* is observable in everyday life as "milk spills but doesn't unspill, eggs splatter but do not unsplatter, waves break but do not unbreak, we always grow older, never younger" (Lebowitz 2007). Hence, all events evolve in such a way that the result is more randomness and equipartition as long as no forces from the outside prevent that (Kümmel 2011, p.131). In other words, the increase of entropy in nonequilibrium processes determines the *thermodynamic arrow of time* which itself shapes the *psychological arrow of time*.[1]

### *3.2.3 The Economy as an Energy-Dissipative/Exergy-Degrading System*

Mainstream economy as a perpetual motion machine of the first kind

All the different growth models presented in Chapter 2 can be sketched by the same Figure 3.3a in which energy is completely absent (recall that the word energy is nearly inexistent in all textbooks discussing economic growth theories). In this conventional representation of society, households provide routine labor and human capital to firms in exchange for wages and capital rents (factor payments). An intermediation sector (banks) enhances the transformation of households' savings into financial capital. The accumulation of past production (of goods and services) and the availability of private and public investments allow firms to invest in physical capital. Firms then combine the different factors of production (physical capital, routine labor, and human capital) to produce goods and services in return for consumption expenditure. Considering the brief introduction to thermodynamics given just above in Section 3.2.2, it is clear that:

> *In standard economics the economy is a closed system in which cycles can occur indefinitely without the need for any energy input. Hence, in this paradigm the economic system is a perpetual motion machine of the first kind; that is, a conceptual artifact that cannot possibly exist in the real world.*

---

could expect that "a closed universe would suffer "heat death" in some distant future, when it will have finally reached equilibrium. Then all radiation energy would be at constant temperature everywhere, and matter would be spread as evenly as possible in space and over all states of motion. All the energy of the universe would have become degraded to thermal energy. Nothing could happen anymore. But who knows whether our universe is closed or open".

[1] Among many others, Hawking (1988) and Zeh (2007) discuss the interrelations between the thermodynamic, psychological, and cosmological arrow of time, which for the latter (currently) points in the direction of the universe expanding.

Enclosing the economy in the surrounding environment

In the 1970s several scientists harshly criticized this nonsense of standard economics, which ultimately led to the establishment of a new paradigm called *biophysical economics*[1] or *thermoeconomics*. In particular, Odum (1971) was one of the first to observe that energy forms differ in terms of quality, as proxied by their exergy content, and that consequently societies with access to higher-quality fuels have an economic advantage over those with access to lower-quality fuels. Furthermore, Odum (1971) noticed that for each dollar flowing into the economy there is inevitably an energy flow in the opposite direction, but *while money circulates in a closed loop, energy exhibits a unidirectional flow*. This flow starts from low-entropy (ordered) energy extracted from the environment that is used for economic tasks, and finally leaves the economic system as degraded high-entropy (disordered) heat. Moreover, Odum (1971) emphasized the fact that solar radiation, the most essential natural energy flow for life from which all other natural flows are derived (such as water and wind flows and all biogeochemical cycles) has absolutely no associated countercurrent dollar flow. Quite similarly, Daly (1985) claims that the circular flow of exchange value represented by virtual money is coupled and maintained by a physical *throughput* of *matter-energy* (and not just energy as emphasized by Odum 1971), which is absolutely not circular, but rather linear and unidirectional. For its part, the analysis of Georgescu-Roegen (1971) was even more centered on materials than energy availability. Hence for him, the economic process is a material-processing open system that converts high quality (low entropy) raw materials into goods and services, while dissipating and discarding large and growing quantities of low quality (high entropy) heat energy and material wastes. In order to understand how order can be made out of disorder in open thermodynamic systems, Prigogine (1967) developed the concept of *dissipative structure*, which describes systems that convert a part of the available input energy into work in order to *self-organize*, a discovery that won him the Nobel Prize for chemistry in 1977.

As depicted in Figure 3.3b, combining these different analyses leads us to enclose the basic economic system of Figure 3.3a in the broader environmental system and to specify that a unidirectional throughput of material-energy of progressively declining quality is absolutely indispensable to allow for the perpetual cycling of the economic system. Within such a framework, the energy-dissipating/exergy-degrading economic system respects the fundamental laws of thermodynamics which are otherwise violated by the economy-only delusion of standard economics. In reality, the economy converts low entropy/low information materials into very low entropy/very high information products and services on the one hand, and high entropy/low information wastes on the other. This unspontaneous decrease in entropy (increase in order) from raw materials to products and services is only possible because of a much higher entropy production (decrease in order) resulting from the degradation of the exergy embodied in the energy flow extracted from the environment and going through the economic system. Indeed, the second law of thermodynamics stipulates that exergy is degraded through the functioning of the economic system since it is composed of multiple irreversible processes that imply some entropy creation. Energy enters the economy as a high quality (high exergy content) input in the form of direct solar energy (biomass and water/wind flows) or indirect

---

[1] According to Cleveland (1999), it is Lotka (1925) who coined the term *biophysical economics* in his early call for the use of basic biological and physical principles to aid economic analysis.

stored solar energy in the form of fossils fuels, and nuclear energy in post-industrial stages. Those energy forms are ultimately dissipated into a lower-quality (lower exergy content) heat output that potentially contains zero exergy (and thus zero ability to generate work) if its temperature is the same as the broader environment. In this sense energy cannot be a production factor since it is necessarily conserved. Hence:

> *It is the exergy content of energy that constitutes the most fundamental production factor used up in the economic process. As a consequence, economic growth is above all determined by the ability of society to collect high-quality primary energy forms defined by high exergy contents, convert these primary exergy flows into useful exergy services in the form of light, heat, electricity, and mechanical motion, and in doing so discard low quality energy, i.e. low exergy/high entropy, into the broader environment.*

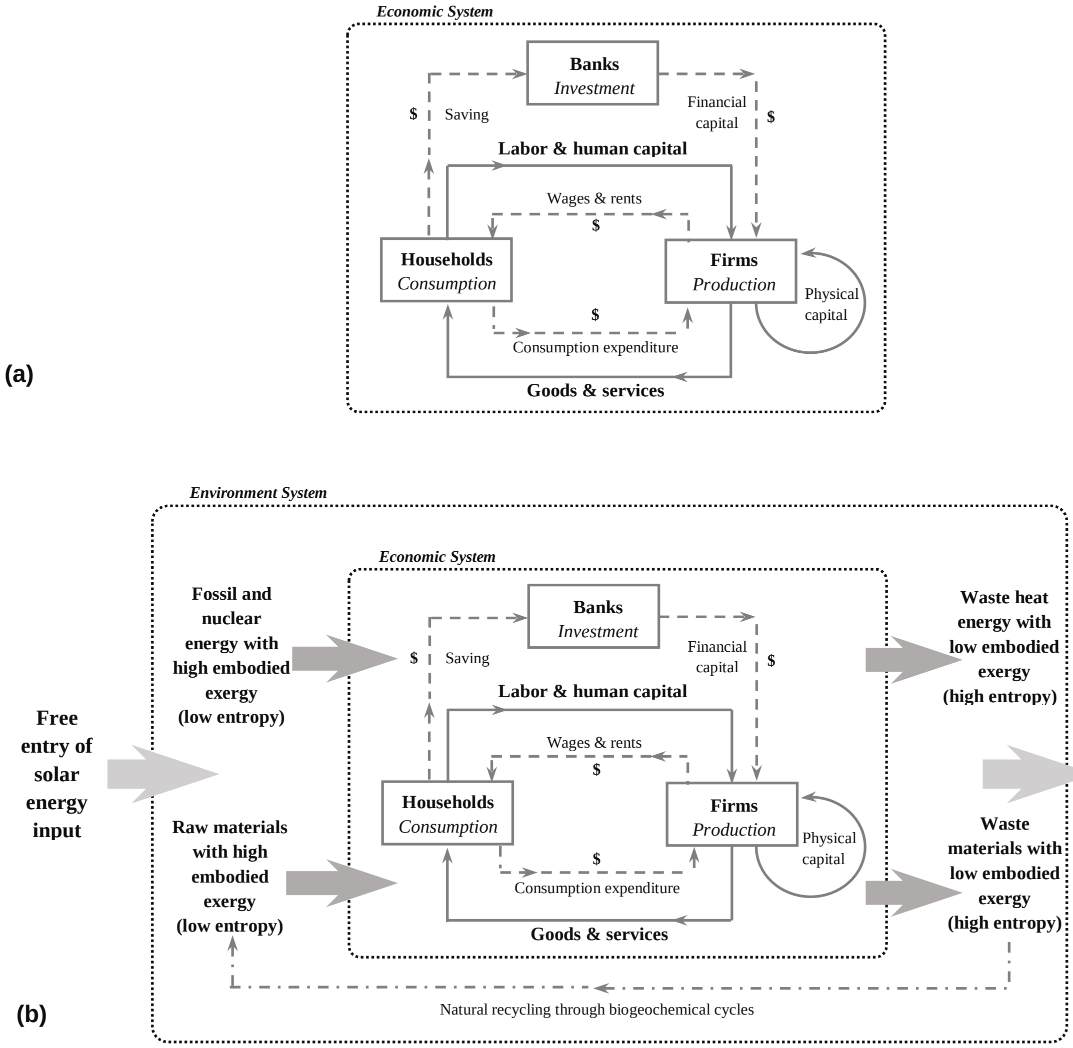


**Figure 3.3 Mainstream vs. Biophysical approaches to the economy-environment system.**
**(a) Mainstream economics view of the economic system as a perpetual motion machine of the first kind. (b) Biophysical economics view of the environment-economy system as a real motion machine. For graphical simplification, the role of the state (or government) as a tax-raiser, investor, and regulator is not shown in this figure.**

### A formal definition for technological change

Following Ayres & Warr (2009), the analytical framework of the biophysical approach is also useful to give a clear and precise definition of technology and technological change, instead of the catch-all solution adopted by standard economics.

Hence:

> *The technological level of the economy represents the efficiency with which primary exergy contained in renewable energy flows (geothermal heat, water, wind, solar, tide and wave) and nonrenewable energy stocks (be it fossil fuels or fissile materials such as uranium) is converted into useful exergy services in the forms of light, heat, electricity, and mechanical power. Thus, the technological level of the economy is the ratio varying between 0 and 1 of extracted primary exergy to consumed useful exergy, and technological progress (respectively regress) reflects an upward (respectively downward) change in the value of this ratio.*

With such a formal definition, aggregate technological level can be computed since primary-to-final and final-to-useful exergy efficiency of all devices of the economic systems can be estimated. Warr et al. (2010) performed such estimation for the UK, USA, Japan, and Austria from 1900 to 2000 and their analyses done in Section 3.3.1 is highly instructive.

Moreover, as a consequence of such a definition of aggregate technological *level*, technological *change* is formally defined as *gains in the aggregate efficiency of primary-to-useful exergy conversion*. An important consequence of such an approach is to consider that the different components of the catch-all definition of mainstream economics (the division and organization of labor, the increasing efficiency of markets, the enhancements of workers' skills, the beneficial effects of inclusive institutions, and the recent contribution of information and communication technologies) could in fact be viewed as *consequences* of technological change and not as technological change itself. Nonetheless, after the accumulation of input factors has been taken into account it would surely be an error to assign all the remaining unexplained growth to the formal definition of technological change given above. First, there is surely a part of the improved economic productivity that comes from the pure *creativity* of the human mind to better organize society without a necessary increase in useful exergy consumption. Second, given the uncertainty in measuring exergy, capital, labor, and GDP quantities (since they are all intellectual constructions emerging from arithmetic conventions), one cannot expect to have a perfect econometric match between historical and reproduced time series of GDP, so that there will always be some apparently unexplained economic growth.

### Entropy production as a measure of pollution

Finally, human activities generate some low exergy (high entropy) wastes in the form of chemical compounds and low-temperature heat. This has been especially true ever since humans started to tap into fossilized solar energy in the form of coal, oil, and gas. Low-temperature heat is rarely damaging for the environment but unfamiliar chemical species that have a chemical potential (measurable by its chemical exergy content) that differs from that of the sinks in which they are stored can cause changes in the delicately balanced biogeochemical cycles linking those sinks. Greenhouse Gas (GHG) emissions resulting from fossil fuel use and other human activities (agriculture mostly) are typically detrimental to the climate balance because they represent chemical wastes that change the entropy content of the atmosphere. Among others, Kümmel (1989) and Ayres (1998) have proposed using entropy disturbance as a measure of pollution but the investigation of this idea is beyond the scope of the present thesis.

It appears crucial now to look at the facts of long-term economic growth through the biophysical paradigm previously presented. Moreover, it will be shown that, far from merely defining material standards of living, systems of energy capture adopted by societies also shape the set of values that people use to interact with each other. Finally, adopting a biophysical approach to the environment-economy system leads to the identification of several limits to substitution between manufactured capital and natural resources. Those limits explain why the so-called de-materialization of the economy is rather unrealistic.

## 3.3 LONG-TERM SOCIETAL EVOLUTION IN A THERMOECONOMIC PERSPECTIVE

### *3.3.1 Exergy Consumption and Societal Development*

#### Useful exergy consumption as the fundamental cause of growth in agrarian economies

In preindustrial agrarian economies, land ownership gave economic and political power to its owners and land was considered as the most crucial factor of production. In fact, this supposed power was due to the photosynthetic collection of solar exergy operated free-of-charge by plants and that people (that are often coerced as slaves or serfs) and animals extracted and transformed into useful exergy services. In such a society, economic growth depends on the capacity to harness increasing primary exergy in the form of food, fodder, motion (from water and wind exergy flows, which are derived from solar exergy), and woodfuel (here again indirect solar exergy), and on the ability to transform those primary exergy resources into useful exergy in the form of light, heat and mechanical power. Land, or rather its three-dimensional extension *space* (i.e. the bio-geosphere), is a production site but not an active factor as long as its capacity to absorb polluting emissions is not binding (Kümmel et al. 2010). In the same way, raw materials remain passive during the production process where “their atoms and electrons are rearranged by capital, labor and energy into the configurations required [to generate] a product or service” (Kümmel et al. 2002). Hence, raw materials do not contribute actively to the generation of value added and can consequently be ignored as long as their finite nature does not constrain growth. Thus, land and raw materials must be seen as *potential constraints* but surely not as *production factors*.

Moreover, in agrarian economies the possibilities of capturing and converting the primary solar exergy flow into useful work is ultimately determined by the forces that organic structures such as animal and human muscles and wood fiber can take and exert. Thus, agrarian economies are by essence Malthusian because of thermodynamic constraints. Heat engines change this situation drastically by allowing people to tap into the huge store of fossilized solar exergy accumulated more than 200 million years ago in the form of coal, oil, and gas.[1] Fossil fuels used in heat engines allow the cheap production of metals from which heat engines and many other machines are built. This positive feedback loop between fossil exergy and raw material extraction greatly expands the amount of accessible natural resources. Most importantly, heat engines convert the chemical exergies of coal, oil, and gas into work *beyond* the limitations of human and animal bodies. Hence, from the first use of heat engines, the level of exergy consumption per capita has been mostly extended through *extrasomatic* exergy, i.e. exergy that is external to the human body, as opposed to *endosomatic* exergy which is derived from food. As put by Kümmel (2011, p.45–46), even if they are obsolete now, the first heat engines based on steam triggered the Industrial Revolution, relieved humans from toil and provided an ever-expanding realm of energy services. To understand the crucial and central role of exergy consumption in the process of societal development, Kümmel (2011, p.16) asserts that human rights, as proclaimed by the Declaration of Independence in 1776, and market economics, as established the same year by *The Wealth of Nations* of Smith (1776), "would not have become ruling principles of free societies had not the steam engines and more advanced heat engines provided the net services that create the preconditions for freedom from toil". A sobering way to understand these assertions is to calculate the number of *energy slaves* in an economy. "This number is given by the average amount of energy fed per day into the energy conversion devices of the economy divided by the human daily work-calorie requirement of 2,500 kcal (equivalent to 2.9 kWh[2] [or 10.5 MJ]) for a very heavy workload. In this sense, an energy slave, via an energy-conversion device, does physical work that is numerically equivalent to that of a hard-laboring human. Dividing the number of energy slaves by the number of people in the economy yields the number of energy slaves per capita". The number of energy slaves at the service of a person has increased in time from 1 throughout the Paleolithic, to roughly 10 in medieval Western Europe, to between 40 and 100 in modern Europe and North America. "And, of course, modern energy slaves work much more efficiently than medieval ones. It is also interesting that Jefferson's original draft of the Declaration of Independence included a denunciation of the slave trade, which was later edited out by Congress. Only after industrialization had provided enough energy slaves could the noble words of the Declaration of Independence be finally put into practice – albeit not without the sufferings of the Civil War".

#### Useful exergy consumption as the fundamental cause of growth in industrial economies

During the last two centuries, animal and human labor has been gradually replaced by exergy-activated machines which have driven down the cost of goods and services (in terms of the number of working hours required to buy such products) and have consequently increased demand and production. In present-day developed countries, this long-term substitution seems to have been the dominant driver of economic growth since the Industrial Revolution (Ayres &

---

[1] Atkins (2010, p.26) gives a complementary explanation based on the molecular interpretation of heat and work given in Section 3.2.1 to elucidate this important aspect of the rise of civilization: that fire from burned biomass preceded the harnessing of fossil fuels to achieve work. He stresses that "the heat of fire–the tumbling out of energy as the chaotic motion of atoms–is easy to contrive for the tumbling is unconstrained. Work is energy tamed, and requires greater sophistication to contrive. Thus humanity stumbled easily on to fire but needed millennia to arrive at the sophistication of the steam engine, the internal combustion engine, and the jet engine."

[2] kWh refers to kilowatt hour, a derived unit of energy equal to 3.6 MJ (megajoule = $10^6$ J).

Warr 2009, p.168). More recently, transistors powered by electricity have started to further reduce biological limitations as they assist the human brain in processing and storing huge quantities of information. Hence, in modern industrialized societies, it is "exergy that drives the machines in mines and on drilling sites, in power stations, factories and office buildings, on rails, road and farms, in the air, and on the sea. In short, it activates the wealth-creating production process of industrial economies" (Kümmel 2011, p.37). The biophysical perspective and the crucial role of exergy consumption are supported by different econometric studies presented in Appendix D for the sake of space. In particular Kümmel has developed a production function called *Linex* which depends linearly on primary exergy and exponentially on quotients of capital, labor, and primary exergy. A particularity of the Linex production function is that it has *variable elasticities of substitution* (VES) between inputs instead of constant ones as in neoclassical production functions (Leontief, Cobb-Douglas, and the more general CES form). There is a good deal of logic in postulating that elasticities of substitution between inputs are not constant over time since "energy and capital are clearly complements during the early stages of economic development (as heavy industry and infrastructure are built up), whereas they become more and more like substitutes later" (Ayres 2008). Ayres (2008) and Ayres & Warr (2009) do not use the Linex production function but directly incorporate useful exergy services (and not primary exergy) in a Cobb-Douglas production function. All these different studies, that alternatively focus on the US, UK, Germany, Japan, and Austria, always deliver the same fundamental result: when useful exergy is incorporated into production functions, the Solow residual (almost completely) disappears, the fit between reproduced and historical GDP is almost perfect (adjusted $R^2$ values are around 0.98), and output elasticities of inputs do not correspond to their cost shares (see Appendix D). Finally, Santos et al. (2016) have recently brought a new perspective to the question of the relative importance of production factors in an attempt to reconcile the biophysical and neoclassical approaches. Focusing on the case of Portugal over the last one hundred years, the worst fits obtained in their analysis all refer to production functions estimated from models where energy is absent from the cointegration space. Hence, the functions that most resemble the neoclassical Cobb-Douglas approach provide the worst fits. On the other hand, they find that the best estimated fit to past economic trends (and lowest TFP component in growth accounting) is a two-input Cobb-Douglas function with quality-adjusted labor and capital, but with capital being a reconstructed variable as a function of useful exergy and labor, and not the historical estimates retrieved from conventional data. In such a case, useful exergy is primordial to defining the actual utilization of capital in production, and estimated values of constant output elasticities for capital and labor are very similar to the average values for historically observed cost shares associated to these factors.

Explaining the economic growth slowdown of the last forty years

Aggregate technological level as defined in Section 3.2.3 as the macroeconomic efficiency of primary-to-useful exergy conversion can be measured since exergy efficiencies of all major exergy conversion processes (and the exergy quantities passing through exergy converting infrastructures) can be estimated. Warr et al. (2010) performed such an estimation for the UK, USA, Japan, and Austria from 1900 to 2000 as shown in Figure 3.4.[1] The aggregate technological level of these four industrialized countries have a similar S-shape form over time, and technological change is formally given by the instantaneous rate of growth of these curves. With such a definition, technological change was rather slow from 1900 to 1945 and then increased considerably up to the 1970s. Since then, gains in the aggregate efficiencies of primary-to-useful exergy conversion have stagnated or declined for all countries.

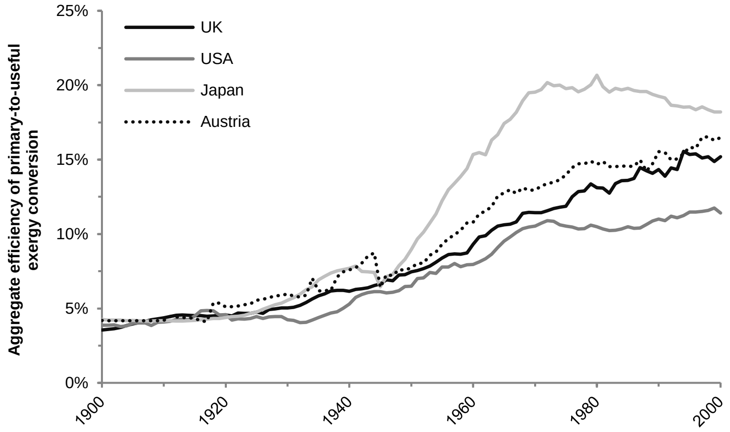


**Figure 3.4 Aggregate technological level of the UK, USA, Japan, and Austria, 1900–2000. Source: Warr et al. (2010).**

It will be recalled that macroeconomic efficiencies of primary-to-useful exergy conversion such as the ones of Figure 3.4 result from the quantity-weighted aggregation of the efficiencies of all primary-to-final exergy converting infrastructures (refineries, power plants, etc.) with the efficiencies of all final-to-useful exergy converting devices (internal combustion engines of cars and trucks, electrical and electronic appliances such as light bulbs, TV sets, and so on). Figure 3.5a shows that the average efficiency of US thermal generation rose from 4% in 1900 to 13.6% by 1925, then almost doubled by 1950, to 23.9%. The nationwide mean surpassed 30% by 1960, but it has stagnated since, never exceeding 33%. From a strict physical point of view, thermal efficiency is of course not equal to a primary-to-final exergy conversion efficiency but it is a good approximation. The important point here is that efficiencies of primary-to-final exergy conversion appear to have stagnated since the 1960–70s in industrialized countries. Regarding final-to-useful exergy efficiencies, here again Figure 3.5b is indicative more than demonstrative. It shows that successive generations of light bulbs have increasing efficiency in terms of lumen emitted per (dissipated) watt, but also that each generation of technology seems to have an intrinsic limit that is ultimately approached. A deeper analysis of the thermodynamic limits of all energy (exergy) converting devices is beyond the scope of this thesis, but it seems clear that if technological change and consequently economic growth have been slowing down since the mid-1970s, it is mostly because of stagnation in the efficiencies of primary-to-final conversion processes, and to a lesser extent because of similar stagnation in final-to-useful exergy conversion efficiencies.

---

[1] Their methodology comprises three distinct stages. The first requires compilation of natural resource exergy, the second is allocation of exergy to each category of useful work and the third is the estimation of the useful work provided by each.

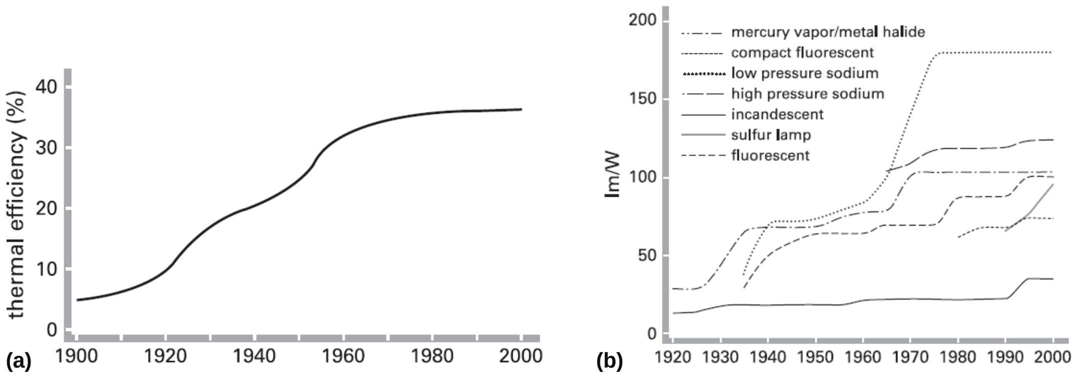


**Figure 3.5 Reaching technological limits in energy efficiencies.**
**(a) US average thermal generation efficiency. (b) Efficiency of electric lighting. Source: Smil (2008, p.237 and 267).**

Furthermore, Figure 3.6 shows that global primary energy consumption per capita increased remarkably from 1945 to the 1970s, but after the second oil crisis, the annual rate of growth of this primary energy input to the economy was rather low.

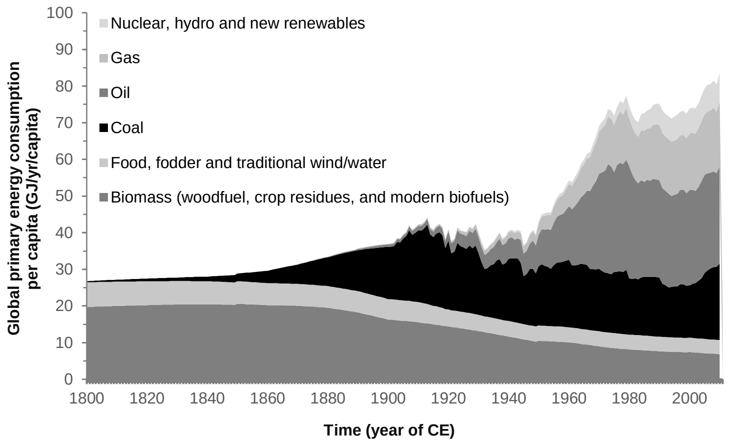


**Figure 3.6 Global primary energy consumption per capita, 1800–2014.**
**Source: energy from Appendix B; population from United Nations (1999, 2015).**

In light of Section 3.2.3, it seems rather clear that the combined patterns of Figure 3.4 (efficiency of primary-to-useful exergy conversion) and Figure 3.6 (primary energy consumption as a proxy for primary exergy extraction from the environment) could explain that growth rates of industrialized economies were significantly higher between the end of World War II and the mid-1970s than in the last forty years (mid-1970s to 2016).

### Law of maximum entropy production for life and human societal evolution

Over its entire time frame (about 150,000 years), societal development has been associated with increasing complexity, order, and information processing over time. Hence, the process of societal development is similar to the evolution of life, which “has tended” during millions of years toward increasing complexity or increasing order with a parallel trend toward increasing information storage and transmission in the form of DNA (or RNA for certain viruses). Both life and societal evolution seem to be subjected to low entropy (or *negentropy*) accumulation, respectively in the biosphere (here defined as the totality of living organisms on Earth) and the technosphere (here defined as the totality of complex human societies); and an associated production of entropy rejected and accumulated in the geosphere (sum of the lithosphere, hydrosphere, cryosphere, and atmosphere) and the broader universe. Lotka (1925, p.357) asserts that evolution of life is a “general scrimmage for available energy”. For Lotka (1922a, p.147), survival is a game governed by the laws of thermodynamics and “in the struggle for existence, the advantage must go to those organisms whose energy-capturing devices are most efficient in directing available energies into channels favorable to the preservation of the species”. Hence, natural selection is driven by the effort of all organisms to try to maximize their total energy throughput, i.e. the energy flux passing through them. If the primary energy supply is limited, the efficiency of the technology used by an organism to capture and convert primary exergy into usable forms is crucial, so that living organisms also tend to maximize their exergy efficiency in order to maximize their energy throughput. In other words, the struggle for survival leads living organisms to self-organize in order to maximize their entropy production. This principle is referred to in the literature as the *law of maximum entropy production* (LMEP, see Swenson 1989; 2009) or *maximum entropy production principle* (MEPP, see Martyushev & Seleznev 2006; 2014). There is currently a great deal of research into determining whether this principle should be considered as the fourth fundamental law of thermodynamics, stipulating that, given their surrounding constraints, all thermodynamic systems in nonequilibrium self-organize to maximize their entropy production. This would be true for living organisms and the process of life evolution (Swenson 2010), Earth systems such as climate (Ozawa et al. 2003; Dyke & Kleidon 2010), and complex human systems and the process of societal evolution (Corning 2002; Raine et al. 2006). More research is clearly needed into that subject, but it seems obvious that the leading nations have always been those that maximize their useful exergy consumption and hence their entropy production. The differential economic growth of nations seems fundamentally to be caused by their relative abilities to consume useful exergy, which depends on the one hand on the (exogenous) favorability of their local environment in terms of available primary exergy, and on the other hand on their (endogenous) capacity to adapt their technology to maximize local extraction and/or imports of primary exergy, and mostly to efficiently convert such primary exergy into useful exergy services. This evolutionary mechanism is so powerful that to maximize the adaptation of technology to the changing environment, human values have also continually adapted to the different systems of energy capture in place.

### *3.3.2 Systems of Energy Capture and Evolution of Human Values*

Taking an evolutionary perspective on the formation of value systems, Morris (2015) deploys a book-length demonstration of the fact that modes of energy capture determine population size and density, which in turn largely determine which forms of social organization work best, and which went on to make certain sets of values more successful and attractive than others. The author starts by identifying the three broad major methods that people have found for capturing energy from the environment, namely as *foragers*, *farmers*, and extractors of fossil fuels (i.e. as *fossil-fuelers*). Then, he observes that foragers overwhelmingly live in small, low-density groups, and generally see political and wealth hierarchies as bad things. By contrast, farmers live in bigger, denser communities, and generally see steep political, wealth, and gender hierarches as fine. Farmers, though, have much less patience than foragers with interpersonal violence, and restrict its range of legitimate uses. Finally, fossil-fuel folks live in even bigger, denser communities, and they tend to see political and gender hierarchy as bad things, and violence as particularly evil. Industrialized societies relying on fossil fuels are generally more tolerant of wealth hierarchies than foragers, although not as tolerant as farmers. Figure 3.7 summarizes the differences between value systems of foragers, farmers, and fossil-fuelers.

For Morris (2015), methods of energy capture largely dictate which demographic regime and forms of organization work best, and these in turn establish what kind of values flourish in society. He summarizes his view in the expression "each age gets the thought it needs". Hence, for him "foraging presented humans with problems that were best solved through shallow hierarchies and abundant violence, which shifted power towards coalitions of losers. Farming created a whole new set of problems, to which hierarchy provided the winning solutions, undermining coalitions of losers. Most of the time, groups that built hierarchical organizations and interpreted justice and fairness to mean that some people (such as men and godlike monarchs) deserved more than others (such as women and peasants) overwhelmed those that did not. . . . In the last two hundred years, fossil fuels have created yet more problems; less hierarchical organizations provided winning solutions to these, and the idea that justice and fairness meant treating everyone more or less the same has largely (but not entirely) swept the field" (*Ibid.*, p.232).

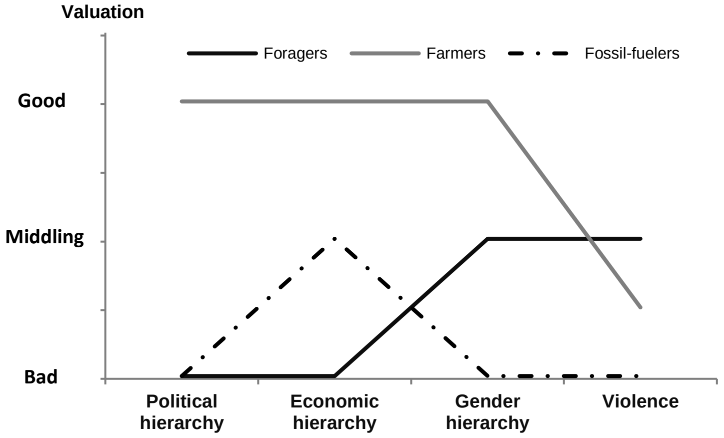


**Figure 3.7 Differences between value systems of foragers, farmers, and fossil-fuelers.**
**Source: reproduced from Morris (2015, p.135).**

Hence, the reason why most foragers share an extremely negative view of political and economic hierarchy, but accept fairly mild forms of gender hierarchy, and recognize that there is a time and a place for violence is that those values are direct consequences of the economic and social constraints created by foraging as a method of energy capture. Indeed, in tiny groups of highly mobile hunter-gatherers, creating and maintaining steep political, economic, or gender hierarchies is very difficult, as is managing relationships without occasional resort to violence.[1]

---

[1] Morris (2015, p. 37–38) adds that another argument for the absence of economic hierarchy within a foraging society is the incapacity to set up a class of rentiers that owns the means of production. "Excluding others from access to wild plants that are scattered over a huge area or wild animals and fish that are constantly on the move is normally impossible. . . . The closest foragers come to owning the means of production is owning man-made devices that improve access to wild food sources, but knowledge of how to make and use [tools] is usually widespread in a foraging group."

Hence, even if the free will of foragers must have pushed some of them to try, say more politically hierarchized but more gender egalitarian and less violent organizations (or any other permutation of these core values), most groups have evolved over time toward the ethical equilibrium described earlier in Figure 3.7, in which values conformed to realities (*Ibid.*, p.43). Similarly, male power over women increased after the agricultural revolution, not because male farmers were more brutish than male foragers, but because this was the most efficient way to organize labor in peasant societies. Hence, if anthropological records contain hardly any examples of farming societies that had no patriarchal organization with gender hierarchy it is because this mode of organization is the most adapted to farming energy capture (*Ibid.*, p.86–87). In the same way, forced (slavery, serfdom) and child labor were widely implemented in farming societies and proved useful for development during thousands of years, but they have been condemned in modern industrialized societies. The reason for that value shift is not that fossil-fuelers became intrinsically more humane than farmers in a few decades. It is rather because, once available, using fossil fuels proved to be more beneficial than coercing people. "And no sooner had free wage labor triumphed than fossil fuels also began dissolving another ancient and indispensable blockage in farming societies' labor markets, the gendered division of labor" (*Ibid.*, p.102). By the same token, Morris (2015, p.242) asserts that if it was not for the tremendous increase in energy per capita consumption brought by fossil fuels, the Enlightenment alone would not have led to a worldwide spread of liberal democracy. For him, it is the explosion of energy after 1800 that made the Enlightenment ideals *viable* all across the West and prevented Northwest Europe and its American colonies from experiencing a conservative reaction against the new ideas, as much of the rest of eighteenth-century Eurasia did. To conclude, "while cultural traditions generate variations on the central themes, energy capture is the motor driving the big pattern" (*Ibid.*, p.10).

### *3.3.3 Limits to Substitution and the Delusion of Dematerialization*

As expressed at more length in Section 3.2.3, all economic systems are open systems that extract and convert low entropy matter-energy into even lower entropy products (and services) and emit high entropy wastes that are freely discarded in the environment. The unspontaneous decrease in entropy associated with the increasing order of matter from raw to refined materials in the form of goods is only possible because an even higher amount of entropy production is associated with the degradation of exergy extracted from the environment. In this complex macro process there are three different limits to substitution between human-made capital and natural capital that explain why the so-called *dematerialization* of society anticipated by many journalists, policy makers, and economists is a rather illusory.[1]

#### Critical natural capital, tipping points, and resilience

---

[1] Ayres & Warr (2009 p.10) observe sarcastically that it is rather misplaced to see economists discussing the currently popular notions of dematerialization and decoupling of economic production from natural inputs since those variables were never coupled in the first place in neoclassical theory.

First, ecological and biophysical economists have suggested that some (if not all) forms of natural capital must have a minimum stock size in order to correctly fulfill their life-support services to the broader ecosystem, and consequently to the economy. Furthermore, scientists have highlighted the existence of non-linear dynamics and feedback control between the different compartments of the global ecosystem. Hence, numerous scientists stress that continuously substituting manufactured capital for natural capital may lead the environment-economy system beyond a threshold (also called *tipping point*) beyond which resilience vanishes. The fear of the existence of such irreversible changes is crystallized by the current research on global climate change and local disruptions of natural ecosystem functions.

### Thermodynamic limits at micro-level define the ultimate technological change at macro-level

Second, at the micro-level, industrial processes have undeniable thermodynamic limits because of the structural properties of materials. Ayres & Warr (2009, p.52–53) highlight that if the different technologies of individual industrial processes depend on design, the possibilities for design depend upon, and are limited by, the specific properties of materials. Some technologies, such as prime movers and many metallurgical reduction and synthesis processes, depend on the temperatures, and in some cases, pressures, achievable in a confined space. These are limited by the strength and corrosion resistance (chemical inertness) of structural materials at high temperatures. In the same way, turbine efficiency also depends on the precision with which blades, piston rings, gears, and bearings can be manufactured, which depends in turn on the properties of the materials being shaped and the properties of the ultra-hard materials used in the cutting and shaping of tools. Ultimately, technological change at the macro level is defined by the limiting efficiency of all metallurgical, chemical, and electronic processes which in turn depend essentially on the properties of structural materials. Evidently, materials have become increasingly specialized over the years. This trend has enabled machines of all kinds to become more efficient and functional. But increased functionality almost always entails more complicated processing and more complex, and costly, capital equipment. The apparent and highly touted trend toward dematerialization is therefore a complete illusion.

### Interdependence between manufactured capital, natural materials, and energy

Third, at the macroeconomic level there is a physical interdependence between (i) manufactured and natural capital on the one hand, and (ii) energy and materials on the other. Both these interdependences place further limits on substitution. Indeed, a classical argument of mainstream economics is to advance the idea of replacing non-renewable natural resource with manufactured capital in order to avoid depletion issues. But one cannot deny that to construct and maintain manufactured capital (tools, machines) requires natural capital and energy, and that additional natural resources are even necessary to support the living expenses (food, water, etc.) of human individuals operating the freshly "replaced" manufactured capital. Hence, "producing more of the substitute, i.e. manufactured capital, requires more of the thing for which it is supposed to substitute" (Stern, 2004, p.43). Moreover, as highlighted in Figure 3.3 of Section 3.2.3, matter and energy are closely interdependent. An illustration of this point is given in Figure 3.8 where the curve $E = f(M)$ is a two-input neoclassical isoquant for a constant level of output $Y$, with $E$ representing energy, and $M$ representing aggregate materials. Producing the input $M$ (materials) involves consuming some energy $E$, and as a consequence it is possible to represent this energy (that should be understood as indirect energy embodied in the final output $Y$) by the function $g(M)$. For simplicity, we (unrealistically) suppose that the extraction of energy from the environment does not necessitate the use of materials. Taking into account the indirect energy embodied in materials results in the *net* isoquant $E = h(M)$.

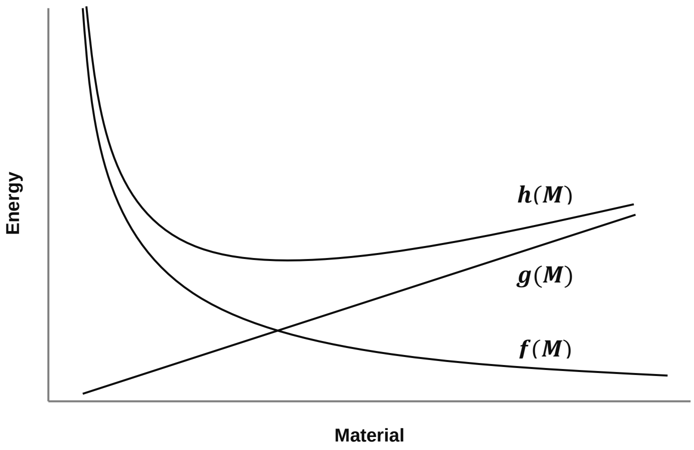


**Figure 3.8 Material-energy interdependence in a two inputs production function.**
**Source: reproduced from Stern (2004).**

To go further, a substitution that appears possible at the micro-level, and is reflected in a micro-level neoclassical production function, does not hold at macro-level. For instance, improving home insulation decreases local heating fuel consumption, but producing the materials needed to insulate homes necessarily requires material and energy; so for the economy as a whole it is by no means obvious that there is any net saving of natural resources. Moreover, if the material costs associated with energy extraction were accounted for in the example in Figure 3.8, decreasing returns to all factors would eventually appear at the macro-level.

This third chapter started by analyzing the misguided reasons for overlooking natural resources, and in particular energy, in mainstream economics. Then, essential concepts such as exergy and entropy were described to understand that the first and second laws of thermodynamics always apply to the economic system and shape its functioning. Applying these fundamental laws to the economic system demonstrate that mainstream economists see the economy as a perpetual motion machine of the first kind, that is, a machine that performs work though endless cycles without any input of energy. Hence, the mainstream conception of the economic system is a conceptual artifact that can absolutely not exist in the real world. Energy has a small share in total production cost not because it is less important than capital or labor as a production factor, but rather because the biosphere and geosphere generate the physical work that we use abundantly and free-of-charge.

As put by Atkins (2010, p. 22), if the first law of thermodynamics was found to be false, “wealth–and untold benefits to humanity–would accrue to an untold extent”. The second law of thermodynamics is also essential to understand that in reality the economic system is an open system (in the thermodynamic sense) that extracts and converts low entropy matter-energy into even lower entropy products (and services) and rejects high entropy wastes that are freely discarded in the environment. The unspontaneous decrease of entropy associated with the increasing order of matter from raw to refined materials in the forms of goods is only possible because an even higher amount of entropy production is associated with the degradation of exergy extracted from the environment (see Figure 3.3). This biophysical approach shows that only useful exergy consumption can be considered as the fundamental cause of economic growth. The capacity to extract primary exergy from the environment and the ability to convert it with increasing efficiency into useful exergy services (in the form of light, heat, electricity, and mechanical power) seems to be the principal mechanism explaining economic growth.

With such an approach, technological change can be precisely defined as gains in the aggregate efficiency of primary-to-useful exergy conversion. It is then understandable that structural limits of materials will define the ultimate limit of the aggregate technological level of the economy, and this ultimate limit might be closer than expected. In particular, the stagnation since the 1970s of the aggregate efficiency of primary-to-useful exergy conversion in industrialized countries (Figure 3.4) and the slowdown of global primary exergy consumption per capita at the same time (Figure 3.6), seem to be two important combined causes of the economic growth slowdown endured by these same countries for the last forty years.

As put by Sieferle (1997), "universal history can be subdivided into three parts. Each part is characterized by a certain energy system. This energy system establishes the general framework, within which the structures of society, economy, and culture form. Thus, energy is not just one factor acting among many. Rather, it is possible, in principle, to determine the formal basic structures of a society from the pertaining energetic system conditions". Regarding the causal relation between systems of energy capture and value systems adopted by societies, it is important to stress that "it is not that individuals are caused to adopt values by their society's mode of energy capture. Rather, over the course of long stretches of history, and as a result of innumerable social experiments by inventive humans, the societies that are best organized to exploit available modes of energy capture–by their social structures, economic and political institutions, culture and values–will tend to prevail over and displace other societies that are less well organized. Social forms and the associated values that are ill adapted to human survival and comfort, given available technologies, will give way to more effective institutions and values" (Stephen Macedo in the introduction to Morris 2015, p. XIX).

To conclude, the insightful work of Kümmel (2011) is called on once more. In his vivid book, he proposes two laws of economics that are as fundamental for society as the two laws of thermodynamics. I have added below a personal corollary to each of these two laws.

The first law of economics states that:

> *Wealth is allocated on markets, and the legal framework determines the outcome. As a corollary, having exclusive institutions or cultural traits that are detrimental to growth can be potential impediments to growth but having inclusive institutions or cultural traits that are conducive to growth should not be mistaken for fundamental causes of growth as they are only facilitating factors, if not consequences of growth.*

The second law of economics states that:

> *Energy conversion and entropy production determines the growth of wealth. As a corollary, the differential economic growth of nations is fundamentally caused by their relative abilities to consume useful exergy, which depends on the one hand on the (exogenous) favorability of their local environment in terms of available primary exergy, and on the other hand on their (endogenous) capacity to adapt their technology to maximize the extraction of primary exergy and the efficiency with which primary exergy is converted into useful exergy services.*

If this second corollary is true, and it remains to be clearly proven, future economic growth will depend essentially on (i) continued increase in the aggregate efficiency of primary-to-useful exergy conversion, and/or (ii) continued increase in the extraction of available primary exergy resources. The former point has already been discussed but the later needs to be approached in terms of net energy (exergy).

# CHAPTER 4

# ENERGY-RETURN-ON-INVESTMENT (EROI) OF ENERGY SYSTEMS

> *"The future belongs, not to those who have the most, but to those who do the most with what they have."*
>
> Eugene Odum

The application of the entropy concept has been extensively applied to the economic system in the context of debates on sustainability (Faber et al. 1987; O'Connor 1991; Common & Perring 1992; Smulders 1995; Ayres 1998; 1999; Kåberger & Månsson 2001; Krysiak 2006; Annila & Salthe 2009; Herrmann-Pillath 2011; Garrett 2014; 2015). Those approaches are very interesting but they remain rather theoretical and it is hard to believe that a practical indicator based on entropy will ever emerge to efficiently orient public debates and choices. The reason is simply that entropy is an abstract concept that remains unintelligible for most people. Hence, a lot of research has also been concentrated on a far more practical concept called *EROI*, for *energy-return-on-investment*. EROI measures the simple fact that one must first invest some past energy production to build and operate the infrastructures needed to extract primary energy and refine it into useable forms. Intuitively, the amount of energy extracted from a given resource must be higher than the quantity of energy invested to get that energy. In other words the amount of *net energy* (gross energy extracted minus invested energy) extracted from a given energy stock or flow must be positive for these to be truly considered as resources and not sinks of energy. This same idea translates into having an EROI (gross energy extracted divided by invested energy) superior to one. Not all energy resources have the same EROI and this greatly conditioned the economic growth potential of society. Mostly, EROI is a dynamic indicator as it represents the continuous struggle between the physical depletion of the resource and the technological progresses that improves the net energy benefits of this same energy source. Understanding the future net energy dynamics of societies is crucial to anticipate their evolution.

## 4.1 NET ENERGY AND EROI OF ENERGY SYSTEMS

### *4.1.1 From Surplus Energy to Net Energy and EROI*

Surplus energy and net energy: two names for the same concept

In his masterpiece *Energy and Society*, Cottrell (1955) defines a quantity called *surplus energy* as the difference between the energy delivered by a process and the energy invested in this same process. More precisely, Cottrell observes two primordial conditions that bind energy availability and societal development: (i) the investment of a minimum amount of already extracted energy is needed to find and develop additional amounts of energy from the environment, and (ii) a part of the available energy must be employed to protect one's energy flow from others seeking to use it for their own preservation (Cleveland 1999). Following the first condition, Cottrell (1955) believes that the magnitude of the surplus energy delivered by an energy source is its most important feature. To illustrate this argument, Cottrell suggests that it is mainly because fossil fuels deliver greater energy surplus than traditional renewable energy resources (agriculture, woodfuel) that the Industrial Revolution was able to produce unprecedented economic and social growth.[1] Incidentally, changes in the amount of surplus energy delivered to society may be the ultimate limit to economic expansion according to Cottrell (1955, p.31): "It will only be when we get a response from nature, in the form of greatly diminished return in the form of surplus energy, that we can expect the present [industrial] revolution to slow down." Odum (1973) used the name of *net energy* instead of energy surplus to speak about the same notion. For him, "the true value of energy to society is the net energy, which is that after the energy costs of getting and concentrating that energy are subtracted". Hence, it is not sufficient to look at the quantities of gross energy that are available, i.e. stock and flow of primary energies, because the most important variable for any self-organized thermodynamic system (a living organism, an ecosystem, a human society, etc.) is the quantity of energy that is really available once the subsystem of energy extraction has been supplied for its own energy need. Hence, surplus energy or net energy is defined as:

$$Net\ Energy = Gross\ Energy\ produced - Energy\ Invested\ to\ get\ that\ Energy. \qquad (4.1)$$

From a biological perspective, it is indeed logical to think that in order to survive each organism needs to procure at least as much energy as it consumes. For instance, for body maintenance and repair, reproduction, and the raising of offspring, a predator needs to obtain significantly more calories from its prey than it expends catching it. This amount of energy left over after one accounts for the calories used to locate, hunt, kill, refine and utilize the gross energy content of the prey is precisely net energy. It is for this reason that Lotka (1922a; 1922b) has proposed that selective pressure that enhances net energy yields is surely one of the main

---

[1] Accordingly, Cottrell (1955) does not miss that much of what is called technological change is in fact obtained through the use of increasing amounts of higher quality energy (and especially fossil fuels) per laborer to perform a specific economic task. This fact led him to claim that the substitution of human labor by massive quantities of fossil energy was the reason for the Industrial Revolution's success as already observed in Section 3.3.1. But in spite of the major role given to energy, Cottrell did not restrict the source of the human condition to energy availability only. Instead, he claimed that resource availability set the general direction of social change, and that human have in a way, to cope with this constraint (Cleveland 1999).

drivers of biological evolution. Odum (1971) refined this idea by including time and consequently referring to power and not simply energy. For him, only those organisms with sufficient net power gains (i.e. net energy gains per time unit) are able to survive and procreate. He synthetized his view in his *maximum power principle* (MPP): “During self-organization, systems designs develop and prevail that maximize power intake, energy transformation, and those uses that reinforce production and efficiency” (Odum 1995, p.311). Odum proposed that the maximum power principle (MPP) should be elected as the fourth law of thermodynamics but it is clear that the law of maximum entropy production (LMEP) or maximum entropy production principle (MEPP) presented in Section 3.3.1 are more general formulations of the same theory which, as of yet, are not fully accepted by all thermodynamicists.[1] Nevertheless, the fact that some 99% (or may be even more) of all species that have ever lived on Earth are now gone surely means that the “technology” they used to gain an energy surplus became obsolete at some time making them unable to balance energy gains versus losses as their environment evolved (Hall et al. 2009).

### Hierarchy of energy needs

The principle explained above can of course be applied to human systems. The need to generate sufficient net energy to feed themselves, reproduce and adapt to a changing environment will not come to the mind of people in industrialized countries, yet for much of the world’s population today, getting enough food is a crucial matter of concern on a daily basis. We must also put into perspective that the majority of humanity’s history and prehistory (about 2.5 million years) has been spent hunting and gathering to obtain the energy needed to satisfy food and energy supplies (Hall et al. 2009). Technology as we commonly understand it is very recent. Because they were able to generate a sufficient energy surplus efficiently (i.e. generate enough net energy in little time), our ancestors could allocate their remaining time to building shelters, improving camp organization, protecting fellow beings, socializing, child caring and storytelling, and cultural improvement in a broader sense (Hall et al. 2009). Following this idea Lambert et al. (2014) have proposed a hierarchy of these energy needs that is analogous to Maslow (1943)’s pyramid of human needs:

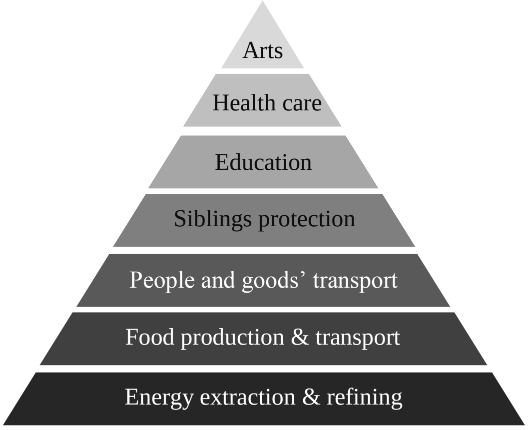


**Figure 4.1 Pyramid of energy needs.**
**Source: adapted from Lambert et al. (2014).**

---

[1] The *Constructal Law* of Bejan (2016) seems to be an even broader generalization of the processes of self-organized designs and evolutions in physical, biological, and social systems.

In this theory, needs perceived as lower in the hierarchy (e.g. extraction and refining of fossil fuels) must be satisfied before needs higher in the hierarchy can become important at a societal level. In other words, energy use for the highest needs, i.e. the performing of arts and other social amenities, are perceived as a societal energy necessity only once all levels beneath them are at least partly fulfilled (Lambert et al. 2014). Furthermore, access to a higher energy need requires the energy source in use to deliver a sufficient net energy surplus. For example, one can only spend energy on health care if there is enough net energy left after lower energy needs (extract, refine and transport energy and food, build shelter, give a minimal education to people) have been fulfilled.

### Net energy increases and human evolution

Key events in human evolution allowed a progressive increase in the amount of net energy extracted from the environment and induced an access to increasingly higher energy needs: (i) the development of spear points and knife blades to exploit a much broader and larger animal-resource base for food and skins; (ii) the recovery of the energy contained in the bonds of the carbon chains of wood and released through combustion; (iii) the concentration of solar energy in photosynthetic plants through agriculture; (iv) the more recent exploitation of wind and water flows; and (v) the later extraction and use of fossil fuels (Hall et al. 2009).[1] By diverting the photosynthetic energy captured from many different species in natural ecosystems to the few cultivars that humans have selected, the development of agriculture (iii) is clearly an improvement in net energy gains. This massive increase in food production per unit of land thanks to agriculture has not led to an increase in food consumption per person but rather to a slight increase in global population and the initial development of cities, bureaucracies, hierarchies, communications and so on. Yet, global human population remained relatively stable for thousands of years and barely grew before 1900. This comes from the fact that during most of human’ history, their energy systems did not generate sufficient levels of energy surplus since humans themselves did most of the physical work as laborers or more sadly as slaves (Hall et al. 2009). Even with draft animals and woodfuel as energy resources, all human societies have prospered for more or less time and then inevitably collapsed (as energy recovery encountered decreasing marginal returns according to Tainter, 1988). Greater energy surpluses were indispensable for humans to more significantly increase their presence on Earth. Fossil energies, and especially oil, are clearly the key energy sources that have allowed a part of humankind to tremendously increase its net energy gains and let its societies enter modernity through the Industrial Revolution.

To ease the description of such evolutions and the ones to come, the concept of energy-return-on-investment, abbreviated to EROI, is more practical than the associated notion of net energy flow previously defined. Contrary to the net energy variable, EROI is a dimensionless ratio which is defined as:

---

[1] As already highlighted in Chapter 3, humans have increased the rate at which they exploit additional energy forms (and also non energy natural resources) through a continuous cumulative process rather than strict transitions from one form of energy or matter to another. Furthermore, this exponential increase in matter-energy consumption has been made towards higher quality (i.e. lower entropy content).

$$EROI = \frac{Gross\ Energy\ Produced}{Energy\ Invested\ to\ get\ that\ Energy}. \tag{4.2}$$

Net energy and EROI are logically related according to:

$$Net\ Energy = Gross\ Energy\ Produced * \left(1 - \frac{1}{EROI}\right). \tag{4.3}$$

EROI is a unitless ratio used to compare outputs to inputs and is therefore more convenient than net energy, which is a finite amount of energy (Murphy & Hall 2011b). It is important to state that the calculation of EROI must use energy inputs and outputs in units assigned to energy carriers such as electricity, gasoline, or steam. An EROI ratio of "20:1" must be read "twenty to one" and implies that a particular process or energy source yields 20 Joules for an investment of 1 Joule, so that the net energy flow is 19 Joules.

Interests of the EROI concept

The EROI of a given energy resource provides multiple useful information:

- First, as a numerical output, EROI allows a comparison between the ability of different types of energy to deliver net energy to society. An energy resource with an EROI inferior to one is in fact a sink of net energy, i.e. it consumes more energy for its use than it delivers to society. The higher the EROI of a given resource, the higher its capacity to deliver net energy to society per unit of invested energy.

- Second, EROI is a measure of the quality of a resource where quality is here defined as the ability of an energy resource to support the economic system. This means in simple terms that when the EROI of a given energy resource is declining over time, a larger part of the society's available energy goes just to get the energy resource instead of being used to run the rest of the economy.

- Third, using EROI measurements in conjunction with standard quantity measures of energy like reserve levels, additional insight is determined about the actual net energy gains of an energy resource. For instance, oil sands of Canada are said to represent 170 bboe (billions barrels of oil equivalent), but with an EROI of 4:1 only three quarters of these 170 bboe will represent net energy to society (Murphy et al. 2011).

### *4.1.2 System Boundaries and Calculation of EROI*

Selecting the appropriate boundary is a crucial step in any EROI analysis. Generally, controversies surrounding EROI analyses between fuels, such as gasoline and corn-based ethanol for instance, are biased most of the time because they do not involve the same boundaries (Murphy et al. 2011). The boundaries of any energy systems vary along three dimensions: output, input, and time.

Output boundary: nature of numerator and range of denominator

Counter-intuitively, the output dimension determines both the *nature* of the numerator and the *range* the denominator of the EROI ratio. This output dimension runs parallel to the energy process chain which consists in: extraction, intermediate processing and end-use. Selecting the level of this dimension is like answering the question “What do we count as energy outputs?”. As said previously, three levels can be basically distinguished in the output dimension. The first is to consider primary energy at the site of extraction, hence the denomination mine-mouth is used for coal, well-head for oil and gas, and farm-gate for biofuels. Since electricity is a final and not primary energy, this first stage is absent for technologies producing renewable electricity from wind, water, and sun. So in the case of fossil fuels and biofuels exclusively, the $EROI_{mm}$ (where “mm” stands for mine-mouth) includes the energy to find and produce the fuel before transport to refining sites:

$$EROI_{mm} = \frac{Primary\ energy\ produced\ at\ the\ mine-mouth}{Energy\ required\ to\ find\ and\ produce\ that\ energy}. \tag{4.4}$$

Even though this is the most commonly calculated EROI for fossil fuels, one would like to know the EROI of gasoline instead of crude oil, or of electricity made from coal instead of coal at the mine-mouth, because those are the final energy forms that are truly used by people. Moreover, renewable energy production only consists in final energy forms so that calculating EROI at the point of use (“pou”) is really important to compare fossil fuels and renewable technologies. In such a case, the $EROI_{pou}$ up to the point of use (pou) where in this case energy inputs (denominator) include the energy to find, produce, refine, and transport energy up to its point of use:

$$EROI_{pou} = \frac{Final\ energy\ delivered\ at\ the\ point\ of\ use}{Energy\ required\ to\ get\ and\ deliver\ that\ energy}. \tag{4.5}$$

Another step is to take into account not only the energy to get and deliver energy up to the point of use, but also the energy invested to actually use this energy. For example, using gasoline in a car requires maintaining general infrastructures such as bridges, highways and the car itself, and the production of all those assets are energy consuming. Hence, the extended (“ext”) EROI is defined as:

$$EROI_{ext} = \frac{Final\ energy\ used\ at\ the\ point\ of\ use}{Energy\ required\ to\ get, deliver\ and\ use\ that\ energy}. \tag{4.6}$$

Quite logically, for a given fuel type the following inequality is always true: $EROI_{mm} > EROI_{pou} > EROI_{ext}$. A graphical representation of these different EROIs is given in Figure 4.2 using the energy circuit language developed by Odum and presented in Appendix E.

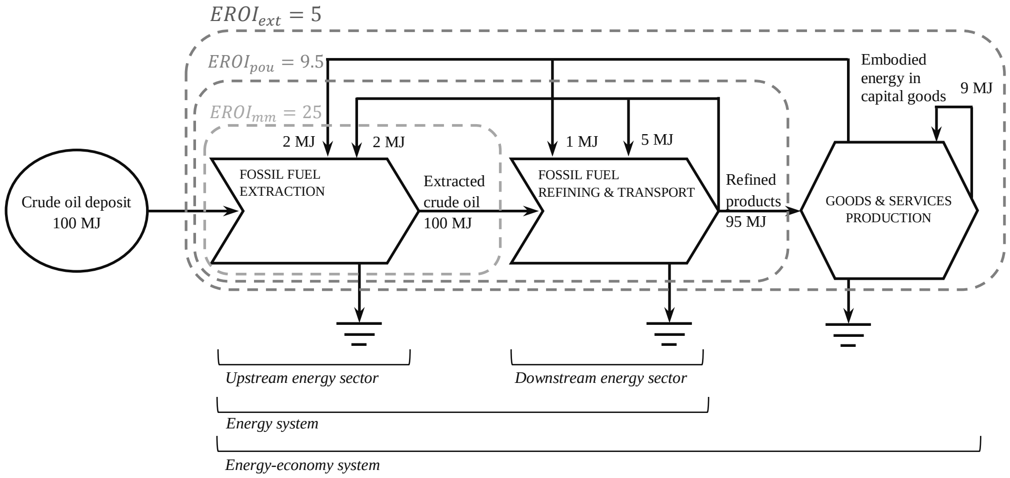


**Figure 4.2 System boundaries of crude oil processing and different EROIs.**

In this example, we focus our attention on the extraction in an upstream sector of a given quantity of crude oil that is then converted by a downstream energy sector in refined product, i.e. diesel, electricity, etc. After accounting for self-consumption by the energy sector, the remaining final energy goes to the rest of the society to produce goods and services. A part of this industrial production is invested in the energy sector in the form of capital. In terms of energy balance, this capital investment represents indirect embodied energy invested in the energy sector and must be accounted for when one calculates an EROI. As an arbitrary postulate, the extraction of 100 MJ of crude oil requires 2 MJ of direct input energy in the form of refined product from the downstream sector, i.e. diesel or electricity, and 2 MJ of indirect embodied energy in capital goods. Thus, the production of 100 MJ at the wellhead has required a total of 4 MJ of energy investment so the $EROI_{mm}$ equals 25:1 (100/4=25). Refining 100 MJ of crude oil requires the investment of 5 MJ of direct energy (electricity for instance) that is assumed to come from self-consumption, and the use of 1 MJ of indirect embodied energy in capital goods. And to deliver the 95 MJ of final energy to society, the former 4 MJ of energy invested in the upstream sector must be accounted for. Thus, in order to deliver 95 MJ of energy to society the total energy investment is 5+1+2+2=10 MJ. Hence, the $EROI_{pou}$ equals 9.5:1 (95/10=9.5). Finally, if one wants to add the 9 MJ of energy needed to build and maintain all the capital infrastructures (roads, bridges, cars, etc.) that enable the use of the 95 MJ of refined fossil fuel products, the $EROI_{ext}$ drops to 5:1 (95/(10+9)=5).

The *range* of the denominators of the three different EROIs in Figure 4.2 is increasing but their *nature* does not change because the same energy inputs are considered in all EROIs: self-consumption, direct external energy, and indirect energy embodied in capital. But other indirect energy inputs could have been included and possibly changed the nature of the denominator from one EROI to another.

### Input boundary: nature of denominator

The other dimension is linked to the denominator of the EROI ratio and establishes the degree of direct and indirect energy and material inputs included in this denominator. Determining this dimension is like answering the question “What do we count as inputs?”. Energy inputs can be classified according to their level of assignation as shown in Figure 4.3.

Level 1 includes only the intrinsic or internal energy of the studied energy source, Level 2 also includes the direct energy input used to extract energy. Level 3 incorporates energy input embodied in materials used to extract the energy such as the energy needed to build drilling rigs, trucks, etc. Levels 4 and 5 respectively include energy embodied in supporting labor and other economic services (Murphy et al. 2011). The choice of this level is of particular relevance because any production process has a great variety of energy inputs. The most obvious is the energy used directly in the process itself, i.e., diesel fuel consumed on a drilling rig (level 2). But one may want to consider as well the energy that has been used to extract and deliver the material inputs to a process, such as the energy used to build the drilling rig or the embodied energy of the helicopter that flies the laborers out to the drilling rig (levels 3). Some components, such as labor, can be considered both direct and indirect energy inputs. Direct labor costs occur as muscle power used on the rig itself while indirect labor costs occur by the energy used to support the paychecks of the workers within steel mills that produce the steel to build the rig. Obviously, the direct labor cost representing direct muscle energy delivered on the drilling rig is almost insignificant compared to the direct energy used by the drilling rig in the form of diesel. Moreover, every production process generates some *externalities* imposed on society and not reflected in the market price of the goods and services. For instance, burning diesel fuel to drill for oil releases sulfur and other greenhouse gas emissions into the atmosphere. This environmental externality is prone to decrease people's welfare because of health issues or global warming. Yet, these side effects are not accounted for when using the heat equivalent of the fuel as the only cost, and thus represent a limitation of current EROI analysis (Murphy et al. 2011).

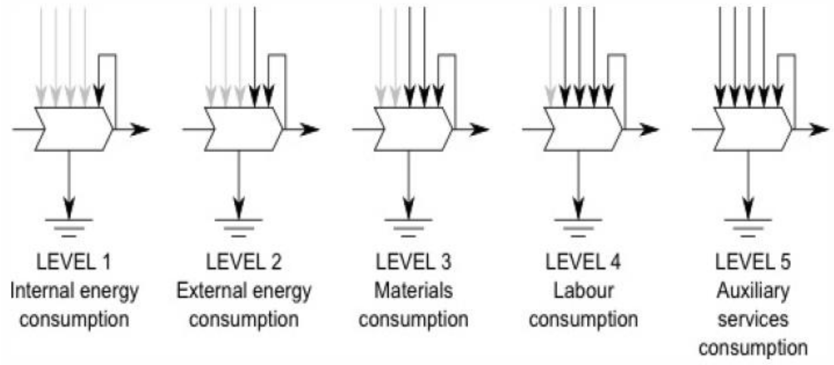


**Figure 4.3 Production process with increasing levels of direct and indirect energy inputs. Source: Murphy et al. (2011).**

Because of industry reluctance to publicly disclose input data, it is often difficult to have explicit energy input, that is why energy inputs have often to be deduced by combining economic inputs and energy intensity factors. Furthermore, capital equipment or material inputs used for energy production consist most of the time in economic data expressed in money invested for a given year. These capital and material inputs have required energy to be produced and transported to their point of use. If they are to be included in EROI analysis as inputs, these economic inputs must be converted into energy inputs in order to account for their embodied energy. The most straightforward approach is to use an energy intensity factor, i.e. a value expressed in units of energy per unit of money like MJ/$ or MJ/€ for instance (Murphy et al.

2011). Considering a given entity, say a national economy or a specific industry sector, energy intensities are easily calculated by dividing the total energy consumption by the total monetary output generated by the specified entity. By multiplying this energy intensity to monetary data (e.g. investment in extraction material such as a drilling rig), one can convert monetary information into energy units and use it in an EROI analysis. The main issue relies in the choice of the energy intensity value. Choosing the basic national-level energy intensity value allows quick conversions but this value averages energy intensities across really different sectors (Murphy et al. 2011). Because oil and gas production is an energy-intensive sector compared to the general economy, industry-specific energy intensity values should be used. As highlighted by Murphy et al. (2011), even though the variability among energy intensities can be considerable, it is still better to choose one of these to convert economic data into energy data rather than omitting these economic data.

Output and input boundaries: a two-dimensional framework

Murphy et al. (2011), in their attempt to build a consistent framework for EROI analysis have represented the information presented so far in a simple, two-dimensional framework as can be seen in Table 4.1. The first dimension of the boundary concerning the energy output of the system, i.e. the numerator of the EROI ratio, is described in the first row from extraction to processing and then end-use. The second dimension of the boundary regarding energy inputs goes along the left side of the table. This Table 4.1 shows that it is possible to have a very tight boundary for the energy output, such as crude oil from an oil well, while having a very wide boundary for energy inputs going from direct energy input to labor used to construct the steel to build the oil drilling rigs ($EROI_{1,aux}$). Of course, doing the opposite is also possible, like calculating the EROI for the gasoline consumed by the end user but only accounting for direct energy used during the entire chain process ($EROI_{3,dir}$). This table is really useful because it establishes a common methodology among researchers that was previously absent. Since most EROI analyses account for both direct and indirect energy and material inputs at the mine-mouth (i.e. well head in the case of oil), Murphy et al. (2011) have defined the $EROI_{1,ind}$ as the *standard* one, so that it can also take the name $EROI_{stnd}$. For practical reasons, the boundary of an EROI analysis will in fact be determined by the data available or the objective of the modeler.

**Table 4.1 Varying EROI from in/output boundaries, adapted from Murphy et al. (2011).**

| | **Boundary for energy Inputs** | **Boundary for energy Outputs** | | |
|---|---|---|---|---|
| | | 1. Extraction (mm) | 2. Processing (pou) | 3. End-use (ext) |
| 1 | Internal energy | $EROI_{1,int}$ | $EROI_{2,int}$ | $EROI_{3,int}$ |
| 2 | Direct external energy | $EROI_{1,dir}$ | $EROI_{2,dir}$ | $EROI_{3,dir}$ |
| 3 | Indirect energy embodied in material inputs | $EROI_{1,ind}$ | $EROI_{2,ind}$ | $EROI_{3,ind}$ |
| 4 | Indirect energy embodied in labor | $EROI_{1,lab}$ | $EROI_{2,lab}$ | $EROI_{3,lab}$ |
| 5 | Auxiliary services and environmental externalities | $EROI_{1,aux}$ | $EROI_{2,aux}$ | $EROI_{3,aux}$ |

As can be observed in Table 4.1, the shaded cells represent EROI with boundaries that favor economic input-output analysis while the other cells favor process-based analysis (also called Life-Cycle Analysis, LCA). In process-based analysis, different kind of energy will be used. For instance, natural gas and natural gas liquids are most of the time associated to crude oil production; in that case the output of the associated EROI has three outputs whereas direct inputs can be as diverse as diesel, electricity, gas, and steam. Thus, defining an EROI for the global combined production of the different outputs will necessitate performing a quality correction of the different energy flows. Indeed, several studies have highlighted the need to perform quality correction of these outputs and inputs since not all joules are created equal, and the economic utility of a unit of electricity is different from the utility of a unit of coal (Kaufmann 1994). The various methods of accounting for quality differences in energy are discussed in Appendix F for the sake of consistency.

Time boundary

Finally, the system boundary may also vary along a third dimension, namely *time*. Indeed, a project can be divided into several steps. As depicted in Figure 4.4, an energy project (like an oil field) first requires an energy input $E_c$ for the construction of the energy facility (drill, pumps, etc.) during time $t_c$. Once the project starts, we call $\dot{E_{out}}$ the annual gross flow of energy and $E_{out}$ the total energy produced over the whole lifetime $t_{op}$. During the same time, an annual input energy flow $\dot{E_{op}}$, is required to operate and maintain the project, resulting in the total energy input $E_{op}$ during time $t_{op}$. At the end of the project lifetime, decommissioning requires the energy cost $E_d$ during time $t_d$ (Murphy et al. 2011).

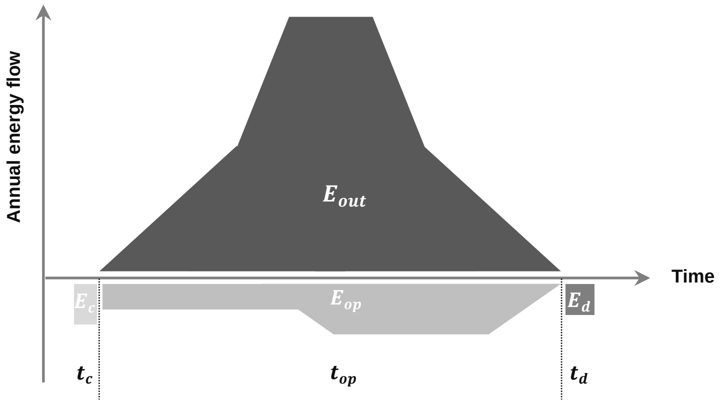


**Figure 4.4 Energy inputs and outputs of a full life-cycle energy project.**
**Source: adapted from Murphy et al. (2011).**

The total net energy output from the project over the whole lifetime is:

$$E_{net} = E_{out} - (E_c + E_{op} + E_d). \quad (4.7)$$

And the EROI from the energy project over the whole lifetime is defined as:

$$EROI_{mm} = \frac{E_{out}}{E_c + E_{op} + E_d}. \quad (4.8)$$

But if one wants to study the dynamic of the EROI of this oil field, annual energy flows must be taken into account to calculate the annual net energy flow:

$$\dot{E_{net}} = \dot{E_{out}} - \left(\dot{E}_c + \dot{E_{op}} + \dot{E}_d\right). \tag{4.9}$$

And the yearly, instantaneous, or more simply annual, EROI is:

$$ER\dot{O}I_{mm} = \frac{\dot{E_{out}}}{\dot{E}_c + \dot{E_{op}} + \dot{E}_d}. \tag{4.10}$$

Considering a large system like the total US oil industry for example, it becomes more difficult to define the lifetime over which energy inputs and outputs are being produced and invested because this system encompasses multiple energy projects with different construction, operation, and decommissioning durations. One must then make the assumption that investments and returns on those investments occur in essentially the same time period (Murphy et al. 2011). Of course, this strong assumption is only accurate if the system is in steady state, i.e. not growing or shrinking. It is important to note that in this case EROI will be reduced in periods of heavy investment (as during periods of high oil prices in the 1970s), and later "inflated" due to the delayed effect of these investments. In order to avoid any confusion between full life cycle EROI and annual (i.e. yearly) EROI, King et al. (2015a) have recently proposed to use more precise names. In their terminology, a *Power Return Ratio (PRR)* refers to an annual EROI because it is a ratio of annual energy flows (energy divided by time, hence "power"), and an *Energy Return Ratio (ERR)* refers to a full life cycle EROI[1] because it is a ratio of total energy quantities.

### *4.1.3 A Dynamic Function of EROI*

Dale et al. (2011) have proposed a dynamic expression of the annual EROI (PRR in the terminology of King et al. 2015a) of a given energy resource as a function of its utilization. Despite the use of such a functional expression of the EROI in a broader theoretical model called GEMBA (Dale et al. 2012), the accuracy of this theoretical model compared to historical EROI estimates of fossil fuels has never been tested. In Section 4.2.2 we provide such global estimations for the EROI of coal, oil, and gas from their respective beginnings of production to present time. In trying to compare these results with the original theoretical model of Dale et al. (2011), we found that it needed to be slightly modified in order to correct two drawbacks.

EROI as a function of the exploited resource ratio

Like Dale et al. (2011) we assume that, for a given year, the annual $EROI_j$ of a given energy resource $j$ (either nonrenewable or renewable) depends on a scaling factor $\varepsilon_j$, which represents the maximum potential EROI value (never formally attained); and on a function $F(\rho_j)$ depending on the exploited resource ratio $0 \leq \rho_j \leq 1$. In the case of nonrenewable energy, $\rho_j(nonrenewable)$ is also known as the normalized cumulated production, i.e. the cumulated production $CumE_{out,j}$ normalized to the size of the Ultimately Recoverable

---

[1] King et al. (2015a) also recommend adding the adjective "external" if the portion of internal energy of the feedstock fuel consumed and dissipated as heat during the process under study (level 1) is not taken into account.

Resource[1], $URR_j$, defined as the total resource that may be recovered at positive net energy yield, i.e. at EROI greater or equal to unity. In the case of renewable energy, $\rho_j(renewable)$ is the ratio of the current production over the Technical Potential[2], $TP_j$, of the resource, defined as the maximum annual flow of the resource that may be recovered at positive net energy yield, i.e. at EROI greater or equal to unity. Hence,

$$\rho_j(nonrenewable) = \frac{CumE_{out,j}}{URR_j} \in [0,1],$$
$$\rho_j(renewable) = \frac{E_{out,j}}{TP_j} \in [0,1]. \tag{4.11}$$

As shown in (4.12), $F(\rho_j)$ is the product of two functions, $G(\rho_j)$ and $H(\rho_j)$. $G(\rho_j)$ is a technological component that increases energy returns as a function of $\rho_j$, which here serves as a proxy measure of experience, i.e. technological learning. $H(\rho_j)$ is a physical component that diminishes energy returns because of a decline in the quality of the resource as $\rho_j$ increases towards 1 (i.e. as the resource is depleted):

$$EROI_j(\rho_j) = \varepsilon_j F(\rho_j) = \varepsilon_j G(\rho_j) H(\rho_j). \tag{4.12}$$

Technological component

In Dale et al. (2011) the technological component $G(\rho_j)$ is a strictly concave function that increases with the exploited resource ratio $\rho_j$. We replace this formulation by a sigmoid increasing functional form (S-shaped curve) that is more in accordance with the historical technological improvements observed by Smil (2005) in the energy industry. Such a formulation is thus convex at the beginning of the resource exploitation, reaches an inflexion point, and then tends asymptotically towards a strictly positive upper limit (Figure 4.5). Hence, our formulation follows the precepts of the original $G_{Dale\ et\ al.(2011)}(\rho_j)$ component of Dale et al. (2011): first, that there is some minimum amount of energy that must be embodied in the energy extraction device; second, that there is a limit to how efficiently a device can extract energy. In other words, we assume that as a technology matures, i.e. as experience is gained, the processes involved become better equipped to use fewer resources (e.g. PV panels become more efficient and less energy intensive to produce; wind turbines become more efficient and increasing size allows exploitation of economies of scale). In our new formulation this

---

[1] According to British Petroleum (2015), the "URR is an estimate of the total amount of a given resource that will ever be recovered and produced. It is a subjective estimate in the face of only partial information. Whilst some consider URR to be fixed by geology and the laws of physics, in practice estimates of URR continue to be increased as knowledge grows, technology advances and economics change. The ultimately recoverable resource is typically broken down into three main categories: cumulative production, discovered reserves and undiscovered resource". On the other hand, Sorrell et al., (2010) highlight that unlike reserves, URR estimates are not dependent on technology assumptions and thus should only be determined by geologic hypotheses. Unfortunately, this apparent contradiction of the URR definition is only a tiny example of the fuzziness of points of view that one could find in the literature regarding the different notions of nonrenewable resources and reserves.

[2] In IIASA (2012), one can read: "Resources of renewable energy are captured by using the concept of technical potential: the degree of use that is possible within thermodynamic, geographical, or technological limitations without a full consideration of economic feasibility."

technological learning is slow at first and must endure a minimum learning time effort before taking off. Moreover, as in Dale et al.'s (2011) original function, our formulation represents the fact that EROI increases from technological improvements are subject to diminishing marginal returns up to a point where processes approach fundamental theoretical limits (such as the Lancaster-Betz limit in the case of wind turbines). In equation (4.13) we have reported the original functional expression found in Dale et al. (2011) that we have called here $G_{Dale\ et\ al.\ 2011}(\rho_j)$ in order to make a distinction with (4.14) that is the function $G(\rho_j)$ that corresponds to the new technological component of the EROI theoretical model.

$$G_{Dale\ et\ al.\ 2011}(\rho_j) = 1 - \Psi_j \exp(-\psi_j \rho_j). \tag{4.13}$$

$$G(\rho_j) = \Psi_j + \frac{1 - \Psi_j}{1 + \exp\left(-\psi_j(\rho_j - \widetilde{\rho_j})\right)}. \tag{4.14}$$

With $0 \leq \Psi_j < 1$ representing the initial normalized EROI with the immature technology used to start the exploitation of the energy source $j$. $\psi_j$ represents the constant rate of technological learning through experience that depends on a number of both social and physical factors that we do not represent. Finally in our new formulation, $\widetilde{\rho_j}$ is the particular exploited resource ratio at which the growth rate of the $G(\rho_j)$ is maximum (i.e. the particular value of $\rho_j$ at which $G(\rho_j)$ presents its inflexion point).

Physical component

The physical resource component of the EROI function, $H(\rho_j)$, is assumed to decrease to an asymptotic limit as cumulated production increases. As advanced by Dale et al. (2011), we follow the argument that on average production first comes from resources that offer the best returns (whether financial or energy) before attention is turned towards resources offering lower returns. Even if this is not completely true at a given moment and for a particular investor, we think that such an aggregated behavior, represented by (4.15), is consistent with long-term economic rationality. A more detailed justification of the decreasing exponential functional form given to $H(\rho_j)$, relying on the probability distribution function of EROI among deposits of the same energy resource is available in Appendix G.

$$H(\rho_j) = \exp(-\varphi_j \rho_j). \tag{4.15}$$

Where $0 < \varphi_j$ represents the constant rate of quality degradation of the energy resource $j$. Furthermore, we correct a failure of the original function of Dale et al. (2011) consisting in the fact that without more specification the asymptotic limit of $H(\rho_j)$ is zero, which implies following equation (4.12) that ultimately energy deposits could be exploited with an EROI inferior to unity (as represented in Figure 4.5). This is in contradiction with the very definition of the URR given previously and with economic rationality. Hence, with the help of the condition found at the end of equation (4.16), we ensure that the EROI ultimately tends towards 1. In order to find this condition, we first consider that $\lim\limits_{\rho_j \to 1} G(\rho_j) = 1$, hence:

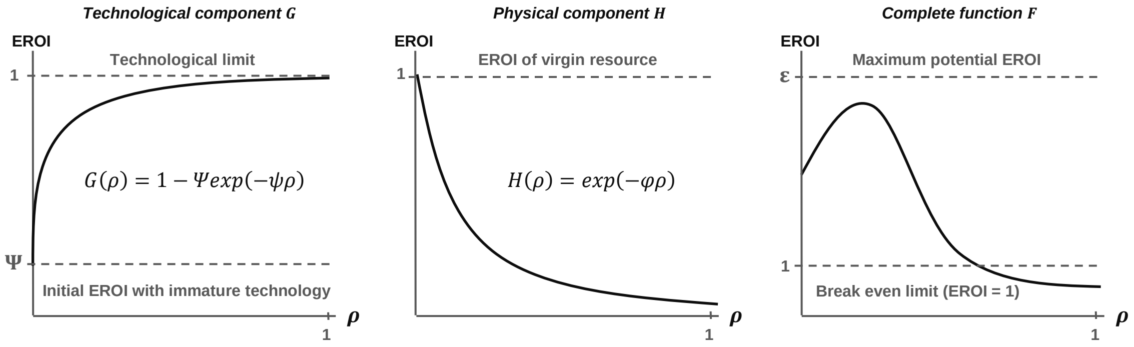
$$\begin{aligned} &\lim_{\rho_j \to 1} EROI_j(\rho_j) = 1 \\ &\Rightarrow \lim_{\rho_j \to 1} \varepsilon_j H(\rho_j) = 1 \\ &\Leftrightarrow \lim_{\rho_j \to 1} \varepsilon_j \, e^{-\varphi_j \rho_j} = 1 \\ &\quad \Rightarrow \varphi_j = \ln(\varepsilon_j). \end{aligned} \tag{4.16}$$

The condition expressed at the end of (4.16) also translates into the fact that there is a strictly positive asymptotic limit $\Phi_j$ to the decreasing function $H(\rho_j)$, as represented in Figure 4.5. The value of $\Phi_j$ is defined as

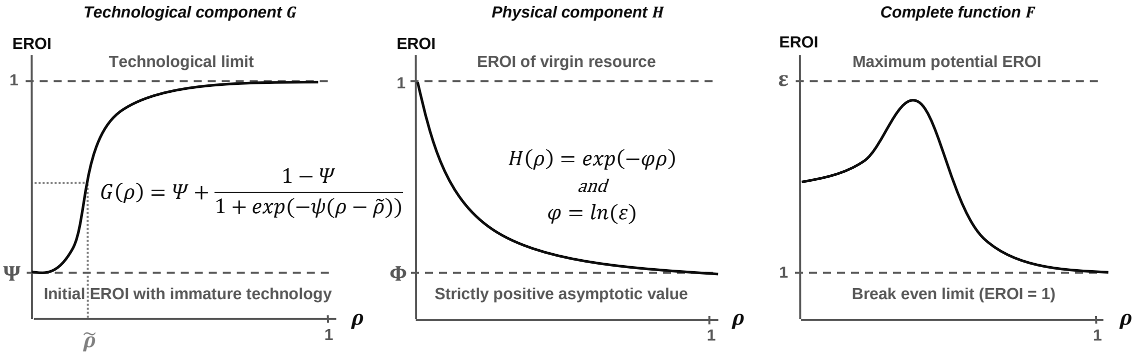
$$\Phi_j = \lim_{\rho_{j \to 1}} H(\rho_j) = \mathrm{e}^{-\varphi_j} = \mathrm{e}^{-\ln \varepsilon_j} = \frac{1}{\varepsilon_j}. \tag{4.17}$$

As shown in Figure 4.5, the amendments operated on the dynamic function of Dale et al. (2011) avoid two drawbacks of the original formulation: (i) the technological learning that serves to increase EROI can now present an increasing S-shape behavior and not a strictly increasing concave form, which is more in accordance with technological diffusion processes; (ii) the exploitation of the energy resource is not possible with an EROI inferior to unity, which was the case with the original function of Dale et al. (2011) and is contrary to economic rationality as it would means that the energy investors invest more energy, and consequently money, than they earn from selling their energy production (even if such an irrational productive behavior might be possible on discrete production sites and for a short time, we postulate that it could not last for long at the aggregated level). However, our new formulation of the theoretical dynamic EROI function makes it more difficult to define the particular value of the exploited resource ratio $\rho_{EROI_j max}$ for which the $EROI_j$ is maximum. This value cannot be found arithmetically anymore (but numerical approximation is of course possible) because of the new functional form we have introduced for the technological component *G*. Nevertheless, the amendments brought to the original theoretical model of Dale et al. (2011) were essential to allow its calibration on the historical price-based estimates of the global EROI of coal, oil, and gas presented in Section 4.2.3.

A very important point that is not stressed in Dale et al. (2011) is that the dynamic function of the EROI does not represent the same physical indicator if one considers a nonrenewable or a renewable energy resource. In the case of a nonrenewable energy resource, equation (4.12) and the right side of Figure 4.5 describe the *average* annual EROI with which the nonrenewable energy is extracted from the environment. But in the case of renewable energy, equation (4.12) and the right side of Figure 4.5 describe the *marginal* annual EROI with which the renewable energy is extracted from the environment. For example, if we take the example of oil for the nonrenewable energy resource, the dynamic EROI function described in this section implicates that the last barrel of oil that will be extracted in the future from the ground will have an EROI just above 1. In the case of a renewable energy resource such as wind, the same function means that the last wind turbine that will be installed, and totally saturate the technical potential of wind energy, will have an EROI just above 1; but of course, in such future situation the whole annual production of energy from wind turbines will have an average EROI far above 1. This difference is off course very important in the context of the energy transition.



**Dale et al. (2011) functional forms for the theoretical EROI model**



**New (present thesis) functional forms for the theoretical EROI model**

**Figure 4.5 Dale et al. (2011) and new functional forms for the theoretical EROI model.**

## 4.2 EROI OF ENERGY SYSTEMS

### *4.2.1 Up-to-Date Representative EROI of Energy Systems*

Despite ongoing debates about the methodology (Zhang & Colosi 2013; Modahl et al. 2013; Brandt, Dale, et al. 2013; King 2014; Arvesen & Hertwich 2015), a large number of studies have been conducted to estimate the EROI of different energy resources. We are not going to present all of them in details, though it would be really important to see differences in methodology and especially the boundaries involved in these different studies. Instead, Table 1 summarized the most "representative EROI" of different energy systems. Of course, there is no such thing as a representative EROI for a given energy type because each energy project has a particular EROI that depends on geographical factors and mostly the input boundary that has been considered to calculate this EROI. The bottom line is that orders of magnitude of energy ratios (be it ERRs or PRRs) are important, precise calculated values are not. The following orders of magnitude can be drawn from Table 4.2. Coal, oil, and gas have respective representative EROIs of about 80–100, 20–30, and 40–60. Hydropower projects have high EROIs of about 50–100:1. New renewable technologies toward which the human future is destined have relatively lower EROIs, especially when their intermittent nature and the subsequent energetic cost for back-up and storage systems is taken into account ("buffering"). Hence, wind power and photovoltaic panels have unbuffered EROIs around 15–20 and 4–12 respectively, but those fall to 4 and 2.5 when the energy cost associated with back-up and storage systems is taken into account. EROIs of first generation biofuels depend heavily on the feedstock. Sugarcane provides ethanol with a relatively high EROI of 10, but all other feedstocks have very low EROI around 1–2, if not under the break-even point of 1.

Furthermore, EROIs are obviously evolving over time as presented in the previous Section 4.1.3. This fact is apparent is apparent from Table 4.2 where EROIs of the same study are given for different time (e.g. oil and gas production in the USA, Canada, and the world). In fact, because of the lack of hindsight concerning renewables and unconventional fuels (such as shale oil, heavy oil, tar sands, shale gas, etc.), EROI trends are especially known for conventional fossil fuels. Time-series analyses have been performed for: global oil (Gagnon et al. 2009), American oil and gas (Cleveland et al. 1984; Hall et al. 1986; Guilford et al. 2011; Moerschbaecher & Day 2011), Canadian oil and gas (Freise 2011; Poisson & Hall 2013), Norwegian oil and gas (Grandell et al. 2011), Mexican oil and gas (Ramirez & Hall 2013), Chinese oil, gas and coal (Hu et al. 2011), American dry gas (Sell et al. 2011) and Canadian dry gas (Freise 2011). All those time series and their analysis are reported in Appendix H. All studies of fossil EROI time-series present the same principal result: declining trends in recent decades with maximum EROI already passed. On the other hand, Table 4.2 shows that so far, unconventional fossil fuels and renewable technologies (except hydro) do not generate as much net energy as conventional fossil energy used to do so. Of course, there are large improvement possibilities for these immature technologies, but for them too, the *First Best Principle* that consists in the use of the best resources first before turning towards lower quality resources applies.

So apparently, all economies will eventually head towards a future where more and more energy is invested in the energy-extraction sub-system of the economy, making net energy delivered to society less available. Before discussing the potential implications of such a future in Section 4.3, it is worth emphasizing that the EROIs of the different fossil energy types used in the economy have never been formally estimated from their respective starting time of production until now. Indeed, all EROI estimates over time have been done in the recent decades (forty years at most) but never from the beginning of fossil fuel extraction in the nineteenth century. In the following section we use a price-based methodology based on King & Hall (2011) to achieve such a goal. Performing such estimations is important to assess how the past evolution of fossil fuel EROIs have influenced the economic growth process.

**Table 4.2 EROI of different energy resources, adapted from Hall et al. (2014).**

| Energy resource | Year | Country | EROI (X:1)* | Reference |
|---|---|---|---|---|
| *Coal* | | | | |
| Coal production | 1950 | USA | 80 | Cleveland et al. (1984) |
| Coal production | 1995 | China | 35 | Hu et al. (2013) |
| Coal production | 2010 | China | 27 | Hu et al. (2013) |
| Electricity production | n/a | n/a | 30 | Weißbach et al. (2013) |
| *Conventional oil and gas (combined production)* | | | | |
| Oil and gas production | 1999 | Global | 35 | Gagnon et al. (2009) |
| Oil and gas production | 2006 | Global | 18 | Gagnon et al. (2009) |
| Oil and gas production | 1955 | USA | 22.5 | Guilford et al. (2011) |
| Oil and gas production | 1970 | USA | 20 | Guilford et al. (2011) |
| Oil and gas production | 2000 | USA | 15 | Guilford et al. (2011) |
| Oil and gas production | 2007 | USA | 11 | Guilford et al. (2011) |
| Oil and gas importation | 2007 | USA | 12 | Guilford et al. (2011) |
| Oil and gas production | 1970 | Canada | 65 | Freise (2011) |
| Oil and gas production | 2010 | Canada | 11 | Poisson & Hall (2013) |
| Oil and gas production | 2008 | Norway | 40 | Grandell (2011) |
| Oil and gas production | 2009 | Mexico | 45 | Ramirez & Hall, 2013 |
| Oil and gas production | 2010 | China | 10 | Hu et al. (2013) |
| *Conventional oil (alone)* | | | | Ramirez & Hall (2013) |
| Oil production | 2008 | Norway | 21 | Grandell (2011) |
| *Conventional dry gas (alone)* | | | | |
| Natural gas production | 2005 | USA | 67 | Sell et al. (2011) |
| Natural gas production | 1993 | Canada | 38 | Freise (2011) |
| Natural gas production | 2000 | Canada | 26 | Freise (2011) |
| Natural gas production | 2009 | Canada | 20 | Freise (2011) |
| Electricity production | n/a | n/a | 28 | Weißbach et al. (2013) |
| *Unconventional fossil fuels* | | | | |
| Deep off-shore oil | 2009 | Gulf of Mexico | 5.5 | Moerschbaecher & Day (2011) |
| Heavy oil | 2005 | California | 5 | Brandt (2011) |
| Tar sands | 2010 | Canada | 6 | Brandt et al. (2013) |
| Tar sands | 2000 | Canada | 4 | Poisson & Hall (2013) |
| Shale oil<br>Shale gas | n/a | n/a | n/a | Despite increasing production in the USA, no actual studies for now. |
| Oil shale | | | | |
| in situ (retorting) technology | 2008 | | 1.8 | Brandt (2008) |
| ex situ (mining) technology | 2009 | | 2.2 | Brandt (2009) |
| *Nuclear* | | | | |
| Electricity production | 2010 | n/a | 75 | Weißbach et al. (2013) |
| *Renewables*** | | | | |
| Hydropower without buffering | n/a | n/a | 50 | Weißbach et al. (2013) |
| Hydropower with buffering | n/a | n/a | 35 | Weißbach et al. (2013) |
| Wind without buffering | n/a | n/a | 20; 16 | Kubiszewski et al. (2010); Weißbach et al. (2013) |
| Wind with buffering | n/a | n/a | 4 | Weißbach et al. (2013) |
| Geothermal (electricity production) | n/a | n/a | 25 | Atlason & Unnthorsson (2014) |
| Wave/Tidal | n/a | n/a | 15 | Halloran (2008) |
| Concentrating Solar Power (CSP) | | | | |
| Parabolic trough without buffering | n/a | n/a | 19 | Weißbach et al. (2013) |
| Parabolic trough with buffering | n/a | n/a | 9 | Weißbach et al. (2013) |
| Fresnel plant without buffering | n/a | n/a | 17 | Weißbach et al. (2013) |
| Fresnel plant with buffering | n/a | n/a | 8.2 | Weißbach et al. (2013) |
| Solar tower | n/a | n/a | 20 | Kreith & Krumdieck (2014) |
| Photovoltaic without buffering | n/a | n/a | 4; 12 | Raugei et al. (2012); Weißbach et al. (2013) |
| Photovoltaic with buffering | n/a | n/a | 2.5 | Weißbach et al. (2013) |
| Biomass (feedstock) | | | | |
| Ethanol (sugarcane) | n/a | n/a | 4; 8 | Soam et al. (2015); Goldemberg et al. (2008) |
| Ethanol (corn or wheat) | n/a | n/a | 0.7; 1.6 | Pimentel & Patzek (2005); Farrell et al. (2006) |
| Ethanol (wood/straw) | n/a | n/a | 2/3.3 | Krumdieck & Page (2013) |
| Ethanol (algae) | n/a | n/a | 0.3; 0.4 | Beal et al. (2012); Seghetta (2014) |
| Biodiesel (soybean or sunflower) | n/a | n/a | 0.6; 5 | Pimentel & Patzek (2005); Pradhan et al. (2011) |

** EROI in excess of 5:1 have been rounded to the nearest whole number.*

*** EROI for renewables are assumed to vary based on geography and climate and are not attributed to a specific region/country.*

### *4.2.2 A Price-Based Estimation of Fossil Fuels Global EROI in the Long-Run*

System boundary

In this section we develop a price-based methodology to estimate the annual (or yearly) EROI of fossil fuels. We used the EROI denomination for convenience but in fact the indicator we measure correspond to a Power Return Ratio in the sense of King et al. (2015a). Regarding the output boundary of our study, it is clear considering our methodology that the different EROI we estimate are all at the mine-mouth or well-head since they concern primary fossil energy. Concerning the input boundary of our study, since we rely on a price-based approach, it makes sense to think that such a price of primary fossil energy covers: direct energy expenditures, indirect energy expenditures from physical capital investment, and indirect energy embodied in workers' supply (i.e. the energy used to provide food, shelter, transport, and all other things consumed by workers) since wages paid to workers in the energy sector are included in the price of the energy produced. As a consequence, if we refer to the nomenclature of Murphy et al. (2011) reproduced in Table 4.1, the different energy (power in the sense of King et al. 2015a) ratios we estimate in this section correspond to "annual $\text{EROI}_{1,\text{labor}}$". Furthermore, the geographic boundary of our analysis is global (i.e. world-wide).

Equations

For a given year, the $EROI_i$ (unitless) of the fossil energy sector, with $i \in (Coal, Oil, Gas)$, can simply be expressed as the ratio of the energy produced $E_{out,i}$ (expressed in exajoule or EJ, representing $10^{18}$ J) to the energy $E_{in,i}$ (EJ) invested in the energy sector *i*:

$$EROI_i = \frac{E_{out,i}}{E_{in,i}}. \tag{4.18}$$

Estimating the *i* different $E_{out,i}$ is rather simple since databases for coal, oil, and gas historical productions are quiet reliable. On the other hand, estimating the quantities of energy $E_{in,i}$ invested in each energy sector is rather difficult and represents the very source of complications when one tries to compute an EROI. Regarding the global economy, it can be proposed that the energy $E_{in,i}$ (EJ) invested in the global energy system *i* corresponds to the quantity of money $M_{in,i}$ (expressed in million 1990 International Geary-Khamis dollars,[1] or M$1990) invested in this sector multiplied by the average energy intensity $EI_i$ (EJ/M$1990) of capital and services installed and used in the energy sector *i* (i.e. the *indirect* quantity of energy consumed by the economic system to generate a unitary dollar consequently spent as capital and services installation and use, plus the *direct* energy consumption of the energy sector *i*). Hence, (4.18) is rearranged as

$$EROI_i = \frac{E_{out,i}}{M_{in,i} EI_i}. \tag{4.19}$$

---

[1] It will be recalled that the 1990 International Geary-Khamis dollar (properly abbreviated Int. G-K. $1990), more commonly known as the international dollar, is a standardized and fictive unit of currency that has the same purchasing power parity as the U.S. dollar had in the United States in 1990.

Of course, the problem now lies in the estimation of the quantity of money $M_{in,i}$ invested in the global energy sector for which very few data exist. Thus, we assume that the unitary price $P_i$ (M\$1990/EJ) of a given energy type divided by the monetary-return-on-investment or $MROI_i$ (unitless) of the energy sector *i* is a proxy for the annual (and not levelized) cost of this same energy. This allows us to estimate the total money $M_{in,i}$ invested in a given energy sector *i* by multiplying the quantity of energy produced $E_{out,i}$ by this sector with the proxy annual cost of this same energy:

$$M_{in,i} = \frac{P_i}{MROI_i} E_{out,i}. \tag{4.20}$$

By injecting (4.20) into (4.19), we obtain that, for each year, the estimated $EROI_i$ at global level is

$$EROI_i = \frac{MROI_i}{P_i \, EI_i}. \tag{4.21}$$

Due to data availability, we have to make two further important assumptions. First, the $MROI_i$ of all *i* energy sectors are the same and correspond to an average MROI of the fossil energy sector. In section 4, we test three different possibilities to estimate this MROI. They deliver very similar results and show that our EROI estimates are almost insensitive to the MROI because the influence of price and energy intensity are far more important. Second, the energy intensities $EI_i$ of all *i* energy sectors are the same and correspond to the average energy intensity *EI* of the global economy. *EI* logically evolves over time and it can easily be calculated for a given year as

$$EI = \frac{\sum_j E_{out,j}}{GWP}, \qquad j \in (Coal, Oil, Gas, Nuclear, All\ renewables) \tag{4.22}$$

where $GWP$ (M\$1990) is the gross world product. In order to calculate the variable *EI*, we have to include the other quantities of energy productions coming from nuclear and renewable energy forms (wind, solar, geothermic, ocean, biofuels, wood, wastes). It follows from these assumptions that (4.21) becomes

$$EROI_i = \frac{MROI}{P_i EI}. \tag{4.23}$$

Then, estimating the global $EROI_{All\ fossil\ fuels}$ of the total primary fossil energy sector is straightforward,

$$EROI_{All\ fossil\ fuels} = \frac{MROI}{P_{All\ fossil\ fuels} EI}. \tag{4.24}$$

Here $P_{All\ fossil\ fuels}$ (M$1990/EJ) represents the average price of fossil energy weighted by the different quantities of produced fossil energies defined by

$$P_{All\ fossil\ fuels} = \sum_i P_i \frac{E_{out,i}}{\sum_i E_{out,i}}. \tag{4.25}$$

The methodology presented above requires having consistent time series for: energy quantities (EJ), energy prices (M$1990/EJ), GWP (M$1990), and an estimation of the $MROI$ (unitless) of the fossil energy sector.

Data

We have used several sources summarized in Table 4.3 in order to estimate the prices of coal, crude oil, and gas. Those were expressed in very different units and we made the accurate conversion so that all prices are expressed in $1990/TJ (here terajoule or TJ, representing $10^{12}$ J, is used instead of exajoule for graphical convenience, see Figure 4.6 and 4.7). Unfortunately, as exposed in Table 1, most of existing long-term time series of energy prices concern American markets. We nevertheless use these data as global proxies by considering that international markets are competitive and that large spreads between regional energy prices cannot last for long due to arbitrage opportunities. This assumption is fairly relevant for oil and gas, especially in the post World War I era. On the other hand, the hypothesis that coal follows a single international price is a rather coarse assumption. Indeed, as coal is really costly to transport, spreads between prices of two different exporting countries have necessarily occurred, especially before 1950. Furthermore, by using a unique price for coal, we do not take into account the manifold qualities of coal (from the high energy content of anthracite to the lowest quality of lignite). As our coal price estimate is representative of anthracite (high quality), our coal EROI is likely a low estimation of the “true” EROI of coal because we surely slightly overestimate the exact quality-weighted global average price of coal. To make things right, we should have computed such a quality-weighted global average price of coal. This would have been possible if we had known both the shares of all the different coal qualities in the total global coal production (i.e. the quality mix of the global coal supply) and their respective prices, for each year between 1800 and 2012. To our knowledge such data is unfortunately not available. In order to express all energy prices in the same convenient unit, i.e. Int.G-K.$1990 per terajoule ($1990/TJ, where 1 TJ = $10^{12}$ J), we have used the US Consumer Price Index found in Officer and Williamson (2016) and different energy conversion factors such as the average energy content of one barrel of crude oil (5.73E-03 TJ), the average energy content of one tonne of coal (29.5E-03 TJ), the average energy content of one thousand cubic feet of gas (1.05E-03 TJ).

**Table 4.3 Sources and original units of recomposed US energy prices, 1800–2012.**

| Energy | Time and spatial coverage | Source | Original unit |
|---|---|---|---|
| Coal | 1800─2012: US average anthracite price. | US Census Bureau (1975a, pp.207–209) from 1800 to 1948; EIA (2012, p.215) from 1949 to 2011; EIA (2013, p.54) for 2012. | Nominal $US/80-lb from 1800 to 1824; then nominal $US/short ton.[1] |
| Oil | 1861─1944: US average;<br>1945─1983: Arabian Light posted at Ras Tanura;<br>1984─2012: Brent dated. | British Petroleum (2015) for the entire period. | Nominal $US/barrel. |
| Gas | 1890─2012: US average price at the wellhead. | US Census Bureau (1975a, pp.582–583) from 1890 to 1915; Manthy (1978, p.111) from 1916 to 1921; EIA (2016, p.145) from 1922 to 2012. | Nominal $US/thousand cubic feet. |

Figure 4.6 presents the different time series of fossil energy prices for coal, oil, and gas expressed in \$1990/TJ. Using (4.25) we have computed from 1800 to 2012 an estimation of the average quantity-weighted price of primary fossil energy shown in Figure 4.7. For this purpose we retrieved primary energy production values (reported in Appendix B) through the online data portal of The Shift Project (2015) which is built on the original work of Etemad & Luciani (1991) for the 1900–1980 time period and EIA (2014) for 1981–2012. Prior to 1900, we have completed the different fossil fuel time series with the original 5-year interval data of Etemad & Luciani (1991) and filled the gaps using linear interpolation. The work of Fernandes et al. (2007) and Smil (2010) were used to retrieve historical global consumption of traditional biomass energy (woodfuel and crop residues).[2]

Total primary energy consumption is also used to compute the average energy intensity of the economy noted *EI* in (4.23) and (4.24). For that we used the gross world product (GWP) estimation of Maddison (2007) from 1800 to 1950 and from the GWP per capita of The Maddison Project (2013) multiplied by the United Nations (2015) estimates of global population from 1950 to 2010. In order to obtain GWP estimates for 2011 and 2012 we used the real GWP growth rate of the World Bank (2016a). Dividing the GWP by the total primary energy production of the world yields the average energy intensity of the global economy presented in Figure 4.8 (expressed here for convenience in MJ per Int. G-K. \$1990). We also present in Figure 4.8 the energy intensity of the global economy over time when traditional biomass energy (woodfuel, crop residues) consumption is not accounted for as seen in some studies (e.g. Rühl et al. 2012). To our mind, not taking into account traditional biomass energy in the calculation of a macroeconomic energy intensity is an important mistake.

---

[1] 1 metric tonne = 1000 kg = 1.10231 short ton; 80-lb = 36.29 kg.

[2] Contrary to popular belief, woodfuel and crop residues still represents 70% of the global renewable energy production nowadays, whereas hydro accounts for 20% and new renewable technologies such as wind power, solar PV, geothermal, and modern biofuels make up the remaining 10%. Furthermore, global historical estimates of traditional biomass energy used in this chapter exclude fodder supplied to draft animals, traditional windmills, and water wheels.

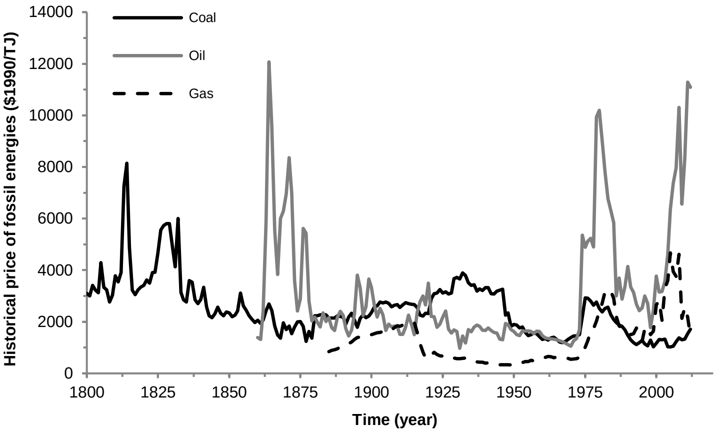


**Figure 4.6 US coal, oil, and gas prices in $1990/TJ, 1800–2012.**

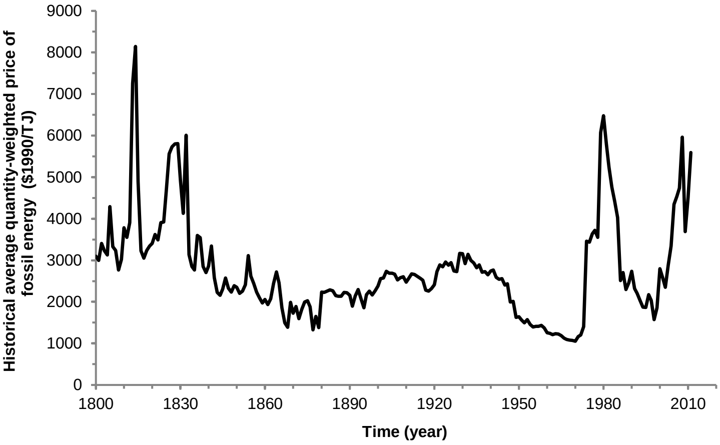


**Figure 4.7 Average quantity-weighted price of primary fossil energy, 1800–2012.**

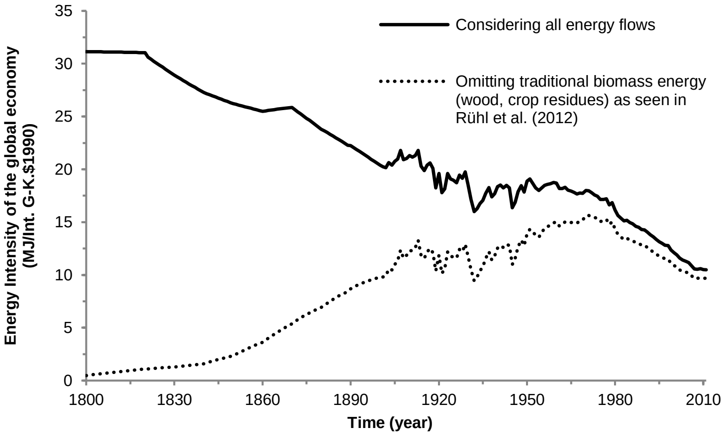


**Figure 4.8 Energy intensity of the global economy, 1800–2012.**

Finally, we follow Damodaran (2015) who claims that the US fossil energy sector MROI roughly follows the US long-term interest rate (US.LTIR retrieved from Officer 2016) with a 10% risk premium. Hence, we compute the MROI of Figure 4.9 following:

$$MROI = 1 + ((US.LTIR + 10)/100) \tag{4.26}$$

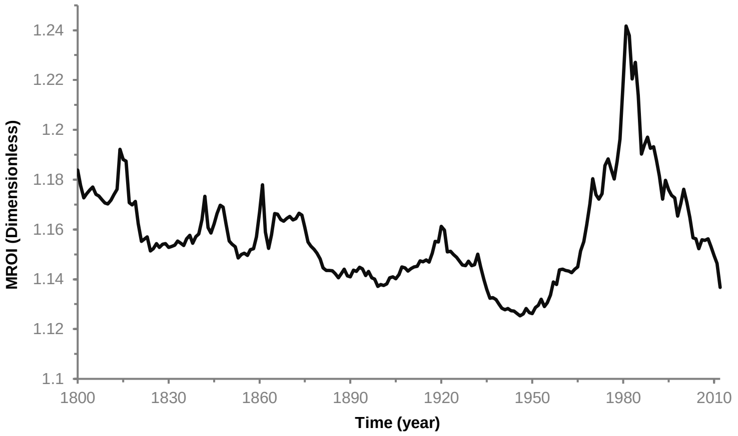


**Figure 4.9 Estimated average annual MROI of US energy sector, 1800–2012. Data source: Officer (2016).**

Results

Figure 4.10 presents the estimations of the global EROI of coal, oil, gas, and of the total primary fossil energy system obtained with the price-based methodology developed above. It is interesting to see that according to our estimations, and contrary to what common sense would suggest, the global EROI of the three fossil fuels (coal, oil, and gas) were not at their maximum in the early years of their respective (reported) productions. Yet, these maximum EROIs seem to have already been achieved in the past for oil and gas global production, with the respective values of 70:1 for oil in 1931 and 207:1 in 1945. EROI of global coal production seems to have broadly steadily increased from 1800 to the present, indicating that maximum EROI has not yet been attained for this energy resource. Furthermore, we can observe in Figure 8 that the global EROI of the total primary fossil energy system has followed the global EROI of coal from 1800 to 1955 and then of oil from 1965 to 2012. From 1955 to 1965, the situation is more difficult to analyze since the EROI of coal and oil are hardly discernable. This is quite logical in the perspective of the historical energy production data reported in Figure 1.9 and Appendix B, where it can be found that 1964 is the precise year during which global oil production exceeded global coal production for the first time.

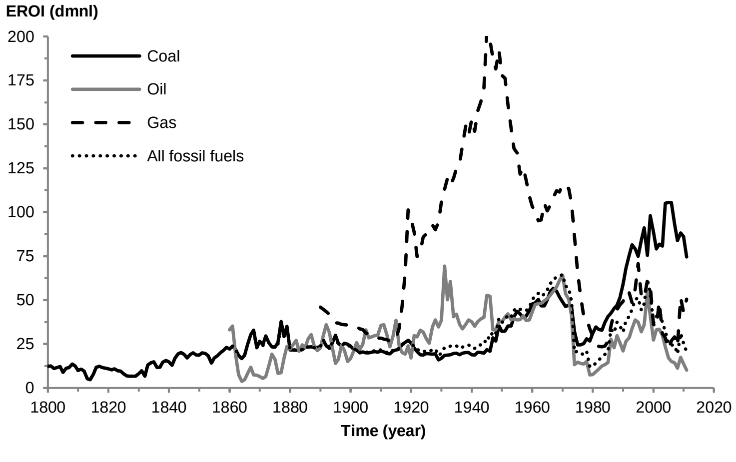


**Figure 4.10 Price-based estimations of fossil fuels EROI, 1800–2012.**

For the sake of space, we provide in Appendix I a sensitivity analysis to MROI and the energy intensity of the economy, and a comparison of our results with existing studies.

### *4.2.3 Price-Based vs. Theoretical Estimations of Historical Fossil Fuels Global EROI*

In order to better analyze the course of the EROI dynamics of Figure 4.10, we are going to compare these price-based EROI estimates to the theoretical dynamic model developed in Section 4.1.3 from the original work of Dale et al. (2011).

Computing theoretical models of fossil fuels global EROI

In order to create such theoretical historical estimates of the global EROI for coal, oil, gas, and total fossil fuels, we first need to determine the respective exploited resource ratios of these resources. Doing so implies defining the Ultimately Recoverable Resource (URR) of each fossil resource. Recall that the URR of a given energy resource is defined as the total energy resource that may be recovered at positive net energy yield, i.e. at EROI greater or equal to unity. These values, presented in Table 4.4, were retrieved from the best estimates of McGlade & Ekins (2015) for oil (Gb: giga barrel), gas (Tcm: terra cubic meters), and coal (Gt: giga tonnes), which for the record are in accordance with the last IIASA Global Energy Assessment report (IIASA 2012). Regarding the coal URR, we find much lower values in other studies, like the average estimate of 1150 Gt (corresponding to 29 500 EJ) given in the literature review of Mohr & Evans (2009). When compared to the order of magnitude of 100 000 EJ found in McGlade & Ekins (2015) and in IIASA (2012), this lower estimation of 29 500 EJ advanced by Mohr & Evans (2009) as an URR corresponds more, according to us, to a proven reserve estimation. However, we will use this lower coal URR estimate to test the sensitivity of our model to this crucial parameter. By combining these URR values with the historical energy production provided in Appendix B, we can compute the exploited resource ratios of the different fossil fuels as defined by equation (4.11). Then, using equations (4.12) and (4.14)-(4.16), we have calibrated the *new* theoretical EROI model on each of the historical estimations obtained with the price-based methodology for coal, oil, gas and total fossil fuels. Best-fit values for parameters $\Psi$, $\psi$, $\overline{\rho}$, and $\varepsilon$ are reported in Table 4.5 and were found using a minimization procedure of the sum of root square errors between the historical estimates of the price-based method and the historical estimates of the theoretical model (value for $\varphi$ is deduced using the final equivalence of relation (4.16)).[1] We have also included the results obtained with a modified version of the original theoretical model of Dale et al. (2011) using equation (4.12), (4.13), (4.15), and (4.16). This *modified Dale et al. (2011) model* consists in taking into account the constraint (4.16), otherwise two problems appear with the purely original model of Dale et al. (2011): (i) the solver was not capable of finding a solution for coal; (ii) the EROI of gas quickly crosses the break-even threshold (i.e. EROI=1:1) after 2033 and then tends towards 0.

---

[1] Robustness of results was tested through a cross validation process: by modifying the data sample (removing some years), parameters of models were re-estimated and proved to remain similar.

**Table 4.4 Global ultimately recoverable resource (URR) estimates for coal, oil, and gas.**
**Source: McGlade & Ekins (2015).**

| **Energy resource** | **Global URR (diverse units)** | **Conversion factors (diverse units)** | **Global URR$^*$ (EJ)** |
|---|---|---|---|
| Coal | 4085 (Gt) | | 105,000 |
| *63% hard coal* | *2565 (Gt)* | *32.5E-9 EJ/tonne* | *83,500* |
| *37% lignite coal* | *1520 (Gt)* | *14.0E-9 EJ/tonne* | *21,500* |
| Oil | 5070 (Gb) | | 31,000 |
| *Conventional oil* | *2615 (Gb)* | *6.1E-9 EJ/barrel* | *16,000* |
| *Unconventional oil* | *2455 (Gb)* | *6.1E-9 EJ/barrel* | *15,000* |
| Gas | 675 (Tcm) | | 27,000 |
| *Conventional gas* | *375 (Tcm)* | *40 EJ/Tcm* | *15,000* |
| *Unconventional gas* | *300 (Tcm)* | *40 EJ/Tcm* | *12,000* |
| Total fossil fuels | | | 163,000 |

$^*$URR values expressed in EJ have been rounded up to the nearest 500.

## Simulation results and comparison with price-based estimates

As could have been expected, the theoretical models provide smooth estimations of historical fossil fuel EROIs. These models also consequently deliver lower values of historical maximum EROIs (i.e. peak EROI) for oil, gas, and total fossil energy. This is summarized in Table 4.6 where we can also see that the historical time of peaking EROIs given by the theoretical models for oil, gas, and total fossil energy are different compared to the ones delivered by the price-based methodology.

Regarding oil, both theoretical models give delayed peaking EROI times compared to the price-based methodology. However, concerning gas and aggregated fossil fuels, peaking EROI times given by the new theoretical model precede the results of the price-based approach, whereas for these same fuels, the modified version of the Dale et al. (2011) model gives slightly lagged (i.e. 1 year) EROI peaking times. Nevertheless, the results of both approaches (price based vs. theoretical dynamic models) are consistent regarding their most important results: the maximum EROI of oil, gas, and total fossil fuels seemed to have already been reached in the past whereas the maximum EROI of coal has not yet been reached.

**Table 4.5 Parameter values of the two EROI theoretical models after calibration.**

| **Model** | **Energy resource** | $\Psi$ | $\psi$ | $\tilde{\rho}$ | $\varepsilon$ | $\varphi = \ln(\varepsilon)$ |
|---|---|---|---|---|---|---|
| New | Coal | 0.0733 | 70.4688 | 0.0471 | 166.2530 | 5.1135 |
| | Oil | 0.0000 | 726.9202 | 0.0004 | 43.7869 | 3.7793 |
| | Gas | 0.1095 | 805.8096 | 0.0025 | 145.2906 | 4.9787 |
| | All fossil fuels | 0.3591 | 247.0671 | 0.0229 | 50.7764 | 3.9274 |
| Modified Dale et al. (2011) | Coal | 0.9844 | 2.0557 | - | 818.2974 | 6.7072 |
| | Oil | 0.5975 | 477.9654 | - | 43.9175 | 3.7823 |
| | Gas | 0.9055 | 350.2591 | - | 145.4809 | 4.9800 |
| | All fossil fuels | 0.7384 | 40.6336 | - | 52.1690 | 3.9545 |

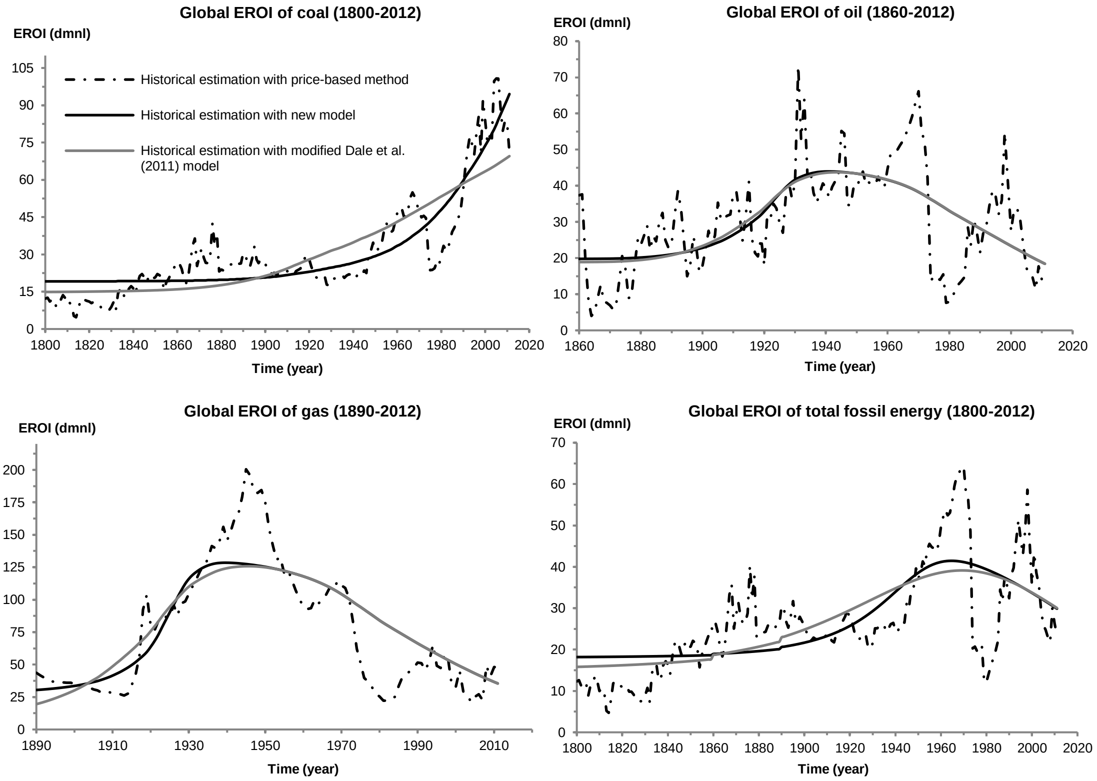


**Figure 4.11 Price-based vs. theoretical estimates of fossil fuels global EROIs, 1800–2012.**

## 4.3 SPECULATIVE FUTURE EROIs AND POTENTIAL IMPLICATIONS

### *4.3.1 Prospective fossil fuel EROIs*

Prospective exploited resource ratios of fossil fuels

Doing some prospective assessments of future global EROIs of fossil fuels is possible by extending the estimations of the theoretical models used previously. For that purpose, we first have to choose hypothetical evolutions for the future exploited resource ratios of fossil fuels. We present such hypothetical evolutions of the exploited resource ratio of coal, oil, gas, and total fossil energy in Figure 4.12. Those were obtained by calibrating increasing sigmoid functions to the historical observed exploited resource ratios.[1] We also propose a deviation range for these prospective exploited resource ratios that corresponds to a change of ten years in their time of maximum growth rate (i.e. from the base prospective exploited resource ratio, we advance or delay the inflexion point of their representative curves by ten years).

---

[1] The exploited resource ratio of a finite resource that necessarily follows a production cycle of Hubbert (1956) type, is quite logically an increasing sigmoid function (i.e. an S-shaped curve). It will be recalled that historical exploited resource ratios are observed but subjected to the hypotheses made on URR values.

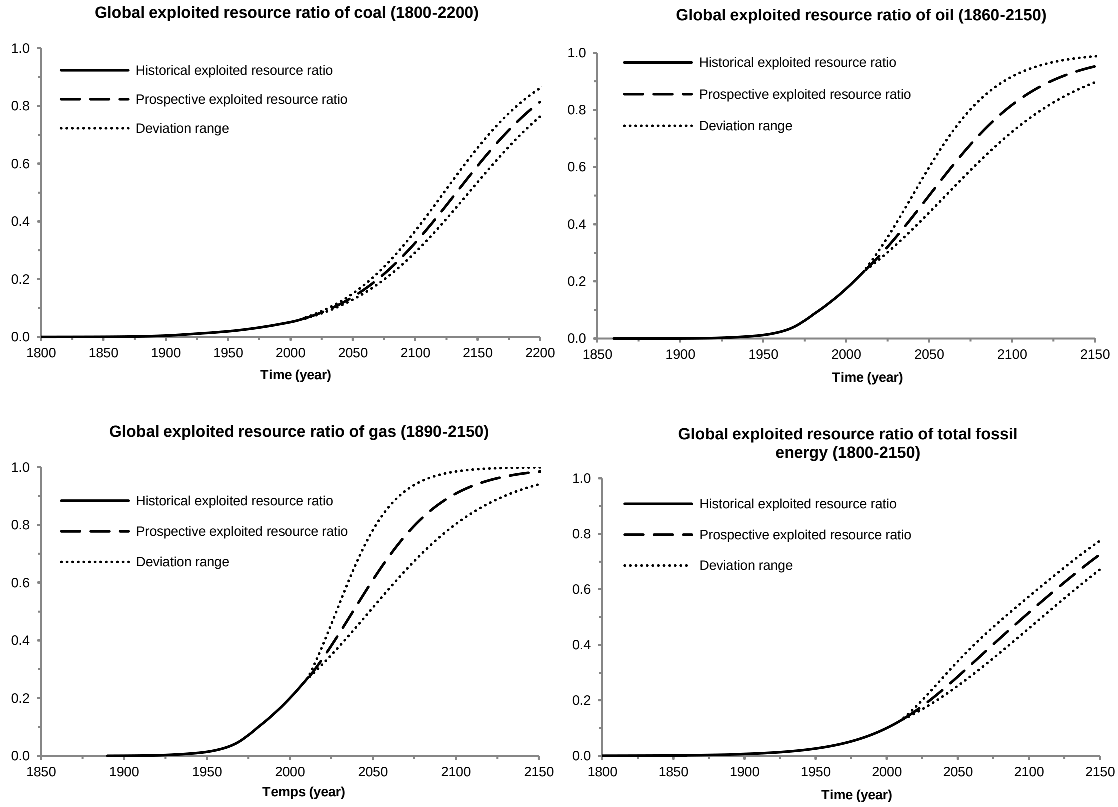


**Figure 4.12 Global exploited resource ratio of fossil fuels, 1800–2200.**
**Hypothetical future exploited resource ratios (dashed lines) are obtained by fitting an increasing sigmoid curve to the historical values (solid lines). Deviation ranges (doted lines) are obtained by advancing or delaying by ten years the time of maximum growth rate (i.e. the inflexion point of the S-shaped curves).**

### Prospective global EROIs of fossil fuels

Based on these prospective exploited resource ratios and keeping the parameter values of Table 4.5, we can obtain prospective EROI for global coal, oil, gas, and total fossil fuels by simply prolonging the theoretical models up to 2150. As shown in Figure 4.13, one of the main results of this prospective exercise is the date and value of the peaking coal EROI that logically differs from one theoretical model to another. With the modified Dale et al. (2011) model, global coal EROI peaks in 2043 at 80:1; whereas with our new formulation of the theoretical EROI model, we estimate that the global coal EROI peak will occur sooner in 2030 but at the higher value of 113:1. Hence, both theoretical EROI models support the idea that, since only 10% of global coal resources have been depleted so far, significant energy gains are still to be expected in the coal sector thanks to coming technological improvements. Furthermore, it is also visible in Figure 11 that changing the exploited resource ratio dynamics, i.e. the production profile dynamics at a given URR, does not change the magnitude of the coal EROI peak but only slightly influences the time of this peak. After its peak, the global EROI of coal decreases in a similar way to other fossil fuels.

Table 4.6 synthetized for the three approaches of this study (the price-based method and the two theoretical EROI models) the time at which the different fossil fuels reach their maximum value and the time at which they cross the particular EROI thresholds of 15:1, 10:1, and 5:1 (the break-even threshold of 1:1 is never formally reached since the constraint (4.16) implies that both theoretical EROI models tend asymptotically towards this value).

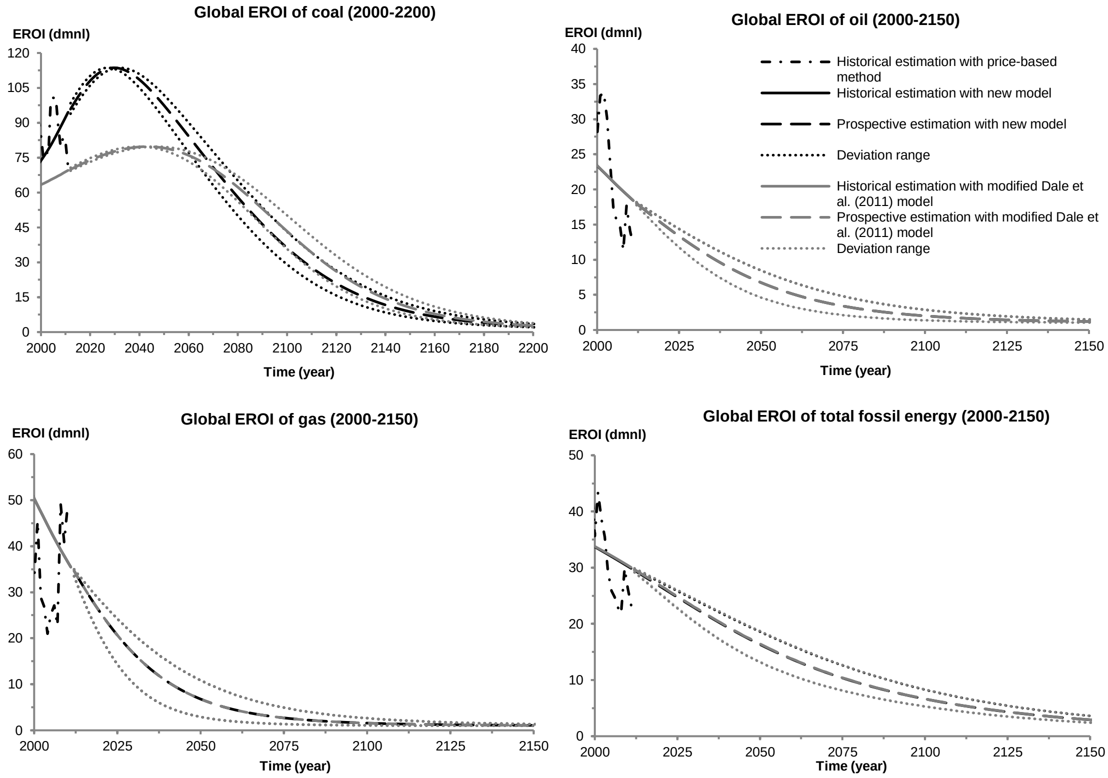


**Figure 4.13 Prospective global EROI of fossil fuels, 2000–2200.**

**Table 4.6 Summarized results of theoretical models and price-based methodology.**

| Energy Resource | Model | Crossing time EROI=15:1 | Crossing time EROI=10:1 | Crossing time EROI=5:1 | Peak EROI time | Peak EROI value |
|---|---|---|---|---|---|---|
| Coal | New theoretical | 2128 | 2143 | 2169 | 2023 | 101:1 |
| | Modified Dale et al. (2011) theoretical | 2140 | 2153 | 2177 | 2043 | 85:1 |
| | Price-based methodology | - | - | - | - | - |
| Oil | New theoretical | 2020 | 2036 | 2060 | 1942 | 42:1 |
| | Modified Dale et al. (2011) theoretical | 2020 | 2036 | 2061 | 1944 | 42:1 |
| | Price-based methodology | - | - | - | 1931 | 70:1 |
| Gas | New theoretical | 2034 | 2043 | 2058 | 1941 | 138:1 |
| | Modified Dale et al. (2011) theoretical | 2034 | 2043 | 2059 | 1945 | 135:1 |
| | Price-based methodology | - | - | - | 1945 | 207:1 |
| All fossil fuels | New theoretical | 2056 | 2078 | 2018 | 1963 | 43:1 |
| | Modified Dale et al. (2011) theoretical | 2057 | 2079 | 2018 | 1975 | 38:1 |
| | Price-based methodology | - | - | - | 1970 | 65:1 |

## Sensitivity of EROI theoretical models to the URR

Given the potentially highly controversial aspect of the prospective results delivered by the theoretical EROI models, sensitivity analysis needs to be carried out. The key parameter of both (*modified Dale et al. (2011)* and *new*) theoretical EROI models is the value retained for the URR. Let us first notice that, as can be seen in Figure 4.14a for the case of coal, dividing the URR by three by assuming an URR of 29 500 EJ (equaling the 1150 Gt best estimate advanced by Mohr & Evans, 2009) instead of the previous 105 000 EJ hypothesis, does not change the estimations of the past theoretical EROI from 1800 to 2012. This is because the curve-fitting procedure (minimization of root square errors sum) generates a new set of constant parameters for which the form of the past coal EROI trend remains consistent. However, as shown in Figure 17b, an URR of 29 500 EJ instead of 105 000 EJ generates a different historical exploited resource ratio that has consequently a different prospective evolution (still approached by a sigmoid increasing function). Finally, Figure 17c shows that the combination of the alternative prospective exploited resource ratio and the new set of constant parameters generate a different prospective EROI that reaches its maximum EROI sooner, 2015 instead of 2023, and at a lower value, 92:1 instead of 101:1. Nevertheless, considering that this sensitivity analysis has consisted in a three-fold division of the coal URR estimation, these results can be considered as quite consistent.

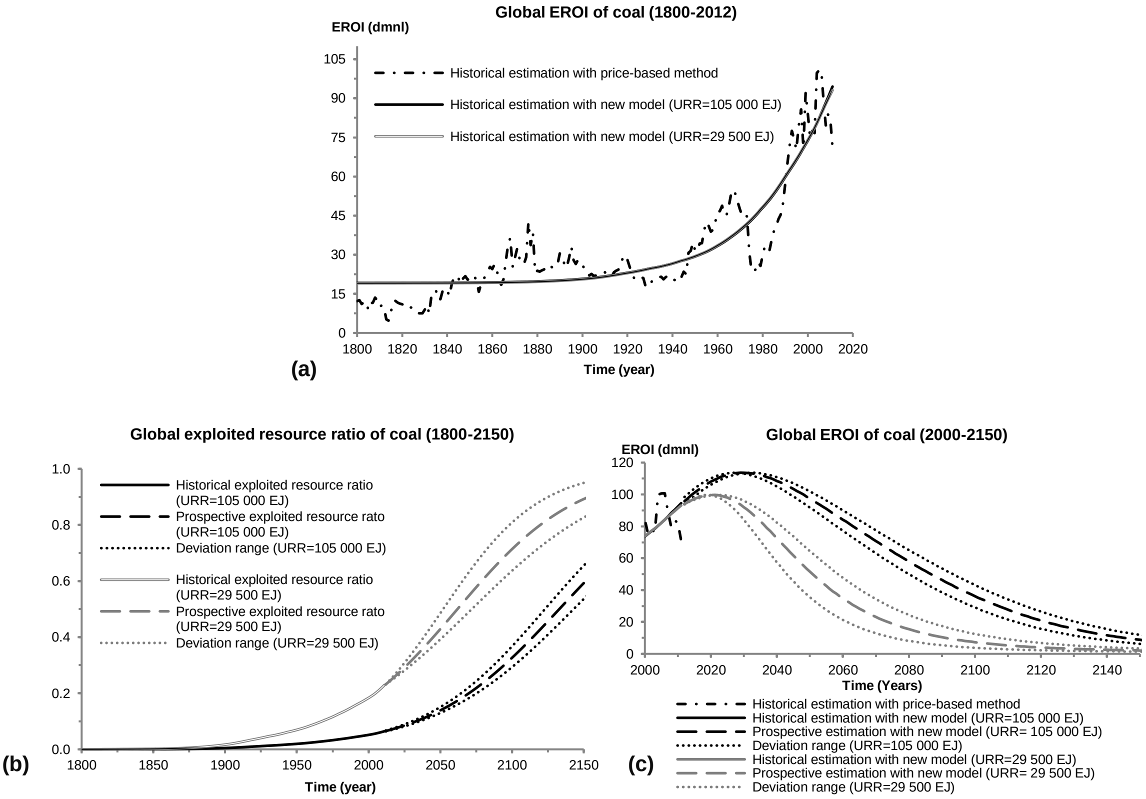


**Figure 4.14 Sensitivity analysis of the theoretical EROI model, 1800–2200.**
**Analysis is performed in the case of coal, using the 29 500 EJ URR estimation of Mohr & Evans (2009) instead of the 105 000 EJ estimate of McGlade & Ekins (2015).**

Furthermore, it is worth stating that if performed on the other two fossil fuels (oil and gas), the sensitivity analysis consisting in a change of their respective URR only changes the slope of their respective decreasing EROI, but under no circumstances could a new EROI peak have been generated. This is mainly due to the fact that by definition in this study, oil and gas comprise both conventional and unconventional fuels since estimations of historical production of unconventional fuels are really scarce. Yet, given the increasing prevalence of unconventional fossil fuels in the primary energy mix, it will be needed to perform again the analyses of the present section in a few decades. This could show that even if it is certain that maximum EROIs have already been reached for conventional oil and gas production, it might not be the case for their unconventional means of production. Indeed, the future preponderance of unconventional fossil fuels production will enable a clear distinction between conventional and unconventional fossil fuel EROIs, which will be of great interests since EROI gains in unconventional production are expected by many whereas our results seems to indicate that the time of increasing EROI has long past for conventional oil and gas production.

If the theoretical model developed in Section 4.1.3 is a good representation of EROI dynamics, Figure 4.13 suggests that the EROI of conventional fossil fuels will decline quite abruptly during the twenty-first century. The EROI of conventional fossil fuels will most surely follow the same path but probably with less amplitude, so that the maximum EROI attained by unconventional oil will probably be lower than the maximum EROI of conventional oil. Of course, if data was available, analysis of regional or national fossil fuel EROIs could be performed and would greatly improve the broad picture that we draw here with a global approach. For now, the so-called renewable technologies cannot deliver as much net energy as fossil energy used to do so since, except for hydro, their EROI is one order of magnitude below that of conventional fossil fuels. It makes no doubt that future technological improvements will improve the EROI of renewable technologies but no one is capable of predicting these future EROIs. Nevertheless, considering that the global remaining hydro potential is far from infinite and will most likely be altered by climate change (Moriarty & Honnery 2012), and adding the intermittent nature of renewable energies and their important metal requirements, it seems rather clear that sooner or later the world economy will have to adjust to low EROIs.

### *4.3.2 Accelerating Depletion of Energy Resources*

Accelerating depletion of energy resources

As showed previously, the EROI of fossil energy is decreasing because we are moving from easily usable resources to deeper, more remote and difficult resources to extract. This basically means that more energy is required to get energy and by consequence less net energy is provided to society. We can picture what is happening by comparing two hypothetical societies. Society A extracts energy with an EROI of 20, while society B does the same thing with an EROI of 5. Society A and B could be two separate societies or the same one at different periods. Let us first consider that both societies extract 100 units of gross energy. Due to a difference in EROI, when society A only needs to invest 5 units of energy to maintain energy investments, society B needs to invest 20 units of energy. As a result, 95 units of net energy are provided to society A and can be used in whatever economic processes to create goods and services, whereas society B only gets 80 units of net energy and consequently generates less economic production (Figure 4.15a). Another way to analyze things is to calculate how much gross energy has to be extracted to maintain a constant supply of 100 units of net energy to society. With an EROI of 20, society A would need to extract 105 units of gross energy, from which 5 would in fact be used as input energy in the extraction process. In society B however, 125 units of gross energy would have to be extracted in order to supply 100 units of net energy (Figure 4.15b).

This simple example shows that declining EROI exacerbates resource depletion by requiring an increase in the extraction of the energy resource to simply offset the decline in EROI. The more the EROI of a fossil resource decreases, the faster the reserve will be depleted. If in the meantime our consumption of fossil energy continues to increase instead or remaining stable, we will exhaust our remaining fossil energy supplies even faster. In other words it also means that if we want to keep a constant flow of net energy to society new fossil energy sources and renewable sources have to be provided with increasing EROIs just to offset the decrease of the EROI of current fossil energy resources. Unfortunately, it is rather the contrary that is happening. First, fossil fuel discoveries and supplies are increasingly relying on unconventional oil and gas (Chew 2014) and as shown in Section 4.2.1, those have far lower EROIs compared to conventional oil. Second, renewable energy sources cannot (for now at least) deliver as much net energy as fossil energy used to do.

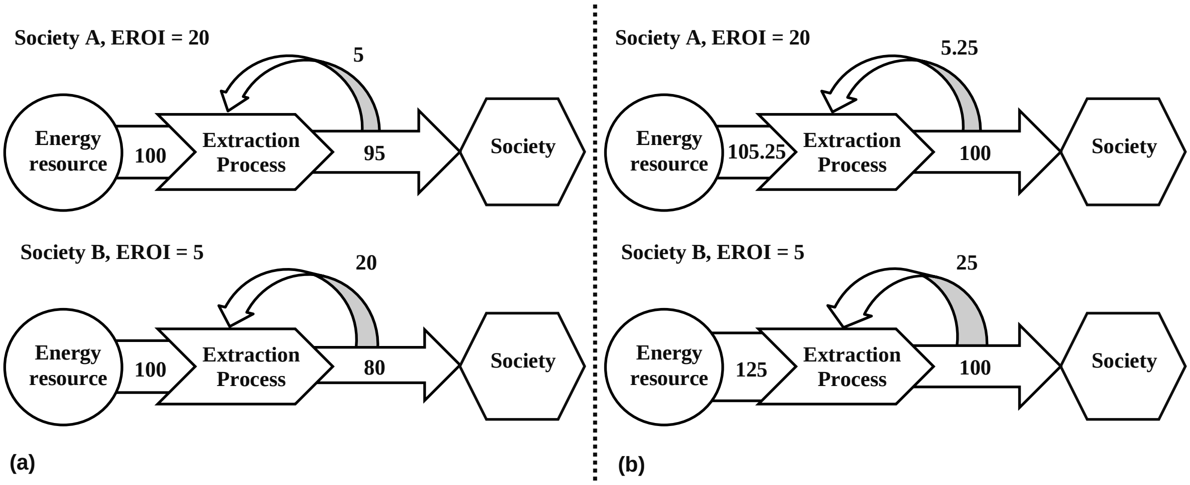


**Figure 4.15 Comparison of two hypothetical societies, A (EROI = 20) and B (EROI = 5). In case (a) both societies extract 100 units of gross energy from the environment; in case (b) both societies extract 100 units of net energy for their functioning. Source: adapted from Murphy & Hall (2010).**

### *4.3.3 The Net Energy Cliff and Higher Energy Prices*

The net energy cliff

If our energy transition away from fossil fuels does not result in their replacement by higher EROI resources, but rather by renewable energies with lower EROIs and their associated intermittent characteristic that generate increasing needs for storage capacity; or if such replacements are possible but delayed too long, then we may be facing a *net energy cliff*. The term net energy cliff was first introduced by E. Mearns who built the graph presented in Figure 4.16, which is derived from equation (4.3) of Section 4.1.1. This graphical representation shows how the relation between the EROI and net energy delivered to society is highly non-linear. If for instance the EROI of a society decreased from 80:1 to 20:1 (a 75% decrease), the part of the gross energy that society disposes off and devotes to gathering energy goes from 1.25% to 5% (i.e. net energy goes from 98.75% to 95% of the gross energy supply). This is a significant change but probably manageable by any society. However, if the EROI drops from 20:1 to 5:1 (a further 75% decrease), now the energy *cannibalized* by the energy sector represents 20% (from a previous stage of 5%) of the gross energy production while only 80% (from a previous stage of 95%) represents net energy available to society. If the EROI of a society declines further below 5, the amount of energy diverted to the energy sector increases exponentially and ultimately almost no net energy remains available to society for doing something different than extracting energy.

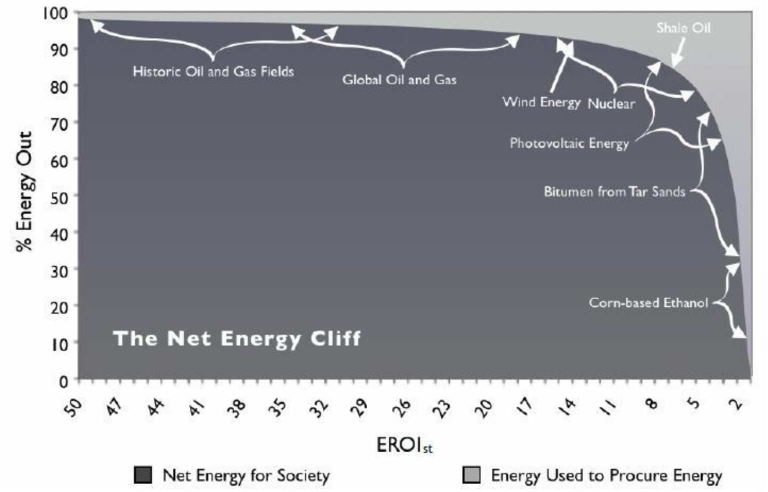


**Figure 4.16 The Net Energy Cliff of E. Mearns.**
**Source: Lambert et al. (2013).**

Presented that way, the importance of net energy to society appears to be obvious; so is the need to correctly anticipate how the EROI of the different energy resources we use will evolve in the future. It is really important to understand how the EROI of the energy resources we use can be increased, what kind of factors can affect their value, etc. Some will argue that technological improvement will allow us to increase the net energy yield of renewable energies and that conversely the same reasoning suggests that technological progress will increase the EROI of unconventional fossil energy resources in the future. For those people, *net energy peak* is no more a problem than *peak oil*. Unfortunately, we will subsequently show in our conclusion that a lot of questions remains on this subject. But before that, let us link the EROI concept with something with which economists are more familiar: the price of energy.

Towards higher energy prices

If fossil fuels EROIs are expected to decline during the next decades as hypothesized in Figure 4.13, the prices of these same energy forms will most likely increase at a rapid pace. Indeed, recall that based on the work of King & Hall (2011) relation (4.21) expresses $EROI$ as a function of the monetary-return-on-investment ($MROI$), the energy intensity $EI$ of the capital invested to get energy (taken as the economy average for simplicity), and the unitary price $P$ of the energy sold on the market. A simple permutation in this expression allows us to isolate price $P$:

$$P = \frac{MROI}{EROI * EI}. \tag{4.27}$$

Figure 4.17 shows graphically how the price of a given energy type evolves as a function of its EROI for several EI values (the MROI is supposed constant and equal to 1.1 for simplicity). Results show that as expected, the relation between energy output price $P$ and the EROI is inverse and highly nonlinear. We can also see that the relation is highly sensitive to the energy intensity of the capital investment. Two other papers from Heun & de Wit (2012) and Herendeen (2015) arrive at the same conclusion: a decreasing EROI, either for a specific energy resource or for the whole economy, implies a highly non-linear increase in energy prices, so that a decrease in EROI from 10 to 4 is not equivalent to a fall in EROI from 100 to 40.

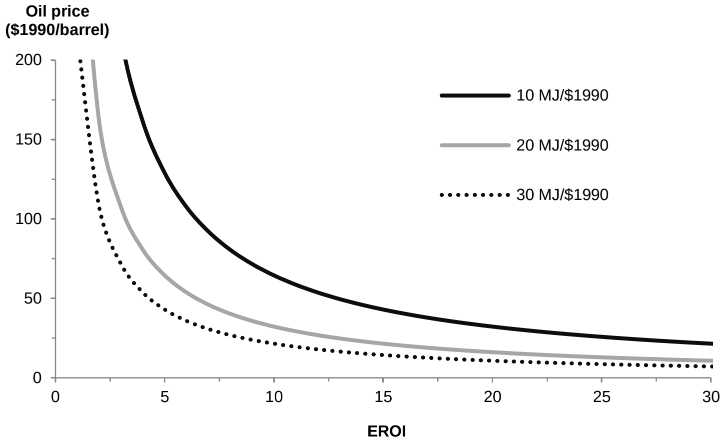


**Figure 4.17 Oil price as a function of EROI and energy intensity of capital investment.**

This fourth chapter showed that the ease of extracting an energy resource, that is to say, its accessibility, is given by its energy-return-on-investment. EROI is a crucial indicator of development because all societies need energy resources that deliver more energy than is invested to use them. Furthermore, it seems logical to think that all different types of societies have a notional minimum EROI required to sustain their level of development. This chapter presented in detail the static (meaning for a given representative year) calculation methodology of the EROI of a given energy system, the different controversies surrounding such a calculation, and hence the limits of this concept. Studies have shown that fossil energy resources to which modern economies have become accustomed and on which they are dependent do not generate as much net energy as before. Indeed, all studies estimating EROI time-series of fossil fuels have so far reached the same result: declining trends in recent decades with maximum EROI already passed. A price-based methodology developed in this same chapter put this issue in a new perspective. It showed that maximum EROI has indeed already been reached at global level for oil and gas production, but for coal, net energy gains are still to be expected thanks to forthcoming technological improvements. On the other hand, recent studies showed that unconventional fossil fuels do not generate as much net energy as conventional fossil energy used to do. Most importantly, renewable technologies in which policy makers and many experts see humanity's future present EROIs that are (currently) lower than past and current fossil fuel EROIs, especially when the intermittent nature of these renewable energy resources is taken into account. Of course, there are is great scope for improvement in these immature technologies, but for them too, the *First Best Principle* that consists in the use of the best resources first before turning towards lower quality resources applies. Hence, all economies will eventually head towards a future in which ever more energy is invested in the energy-extraction sub-system of the economy, making net energy delivered to society less readily available.

# CHAPTER 5

# QUALITATIVE METAL DEPLETION AND EROI

> *"As far as the economic process itself is concerned, we must not ignore the substantial dissipation of matter caused not by purely natural phenomena but by some activities of living creatures, of mankind's, above all."*
>
> Nicholas Georgescu-Roegen

Different studies have highlighted that the requirement of the different sort of metals (i.e. geochemically rare or common) and common minerals (used to produce glass and concrete for example) needed to produce a unit of renewable energy from wind turbines of solar PV is more intense when compared to fossil fuels electricity-producing technologies (Lund 2007; Kleijn et al. 2011; Elshkaki & Graedel 2013; Vidal et al. 2013; Moss et al. 2013). Consequently, an energy transition towards increasing (and possibly complete) renewable energy would require increasing metal consumption (at equivalent energy production). Furthermore, the extraction of metals from deposits, and their concentration in useable forms requires energy. Some early studies have shown that the energy cost associated with metal production increases as metal concentration in deposits decreases (Hall et al. 1986). We can see that a complex interdependence exists between energy and metals sectors and it is the purpose of this chapter to further investigate this relation. Because the energy cost associated with metal extraction is increasing and many different metals are required in renewable technologies, it seems logical to use the EROI concept to link metal and energy sectors. More precisely, in this chapter we will first give an estimation of the current amount of global energy consumed by the metal sector. Unfortunately, doing the opposite calculation of the amounts of the different metals cornered by the energy sector is quite impossible. Then, we are going to build a methodology to investigate how the increasing energy cost associated with metal extraction could influence the EROI of different renewable technologies.

## 5.1 ENERGY AND METAL INTERCONNECTION

### *5.1.1 Basic Interrelation between Energy and Metal Sectors*

Sectors of metal extraction and production represent a significant share of the current total energy consumption. Rankin (2011) estimated that 10% of global primary energy production is consumed by the metal sector. Data from the International Energy Agency (2015) and from Norgate & Jahanshahi (2011) give a lower value of approximately 7%. Table J.1 of Appendix J giving mean energy costs of metal production (Valero & Botero 2002; Rankin 2011; Tharumarajah & Koltun 2011; Ashby 2012) and the quantities of production for different metals in the year 2012 (USGS 2012) confirm the result of Rankin (2011) that at the global scale the metal sector currently requires about 10% of the total primary energy consumption. Of course a degree of uncertainty around these data exists for two reasons: unitary energy costs have different years of estimation; and the method of allocation of the joint cost in case of coproduction with other metals may differ from one study to another.

Conversely, the energy sector consumes a large part of the different metals that are produced across the world. Bihouix & De Guillebon (2010) have evaluated that 5 to 10 % of global steel production is absorbed by the energy sector. It is unfortunately really complicated to give more details about the level of consumption of each metal in the energy sector.

### *5.1.2 Qualitative Aspect of Metal Depletion*

Extracting and refining metals requires energy consumption, so it is easy to define an energy cost of production expressed in GJ per ton of extracted metal. Some numerical data on energy costs are given in Table J.1 of Appendix J, but it would be interesting to assess the evolution of the energy cost associated with metal extraction and production. Such a temporal analysis is important to see how the energy cost associated with metal extraction and production is related to cumulative production.

A first approach consists in analyzing the *Energy Balance Flows* established every year by the International Energy Agency (2015). Such data is presented in Figure 5.1, where the evolution of the final energy consumption of various sectors and the total economy can be compared at the global scale. In Figure 5.1, mining and quarrying activities represent all global upstream activities related to mineral extraction and concentration for both metal and non-metal matter (although fossil fuels like coal are not included). All other final energy consumptions refer to global downstream activities, for either metal refining (of iron and steel, or non-ferrous metals), or non-metallic mineral manufacturing (such as sand, clay, etc.). As can been seen from this figure, the final energy consumption of metal and non-metal refining sectors broadly followed the dynamics of the whole economy at the global scale between 1975 and 2010. However, the final energy consumption of the upstream metal mining sector increased twice as much as the global economy did. Hence, increasing demand from the economy led to an equivalent increasing demand for metal extraction, and a consequent associated final energy consumption (volume effect) in both upstream and downstream metal sectors. However, taking into account that upstream activities, which support metal concentration, are sensitive to ore grade, whereas downstream activities, which represent refining of metals, are insensitive to ore grade, it seems clear that the particular energy consumption pattern of the mining and quarrying global sector is due to the increasing unitary energy cost of metal extraction due to the qualitative depletion of mineral deposits.

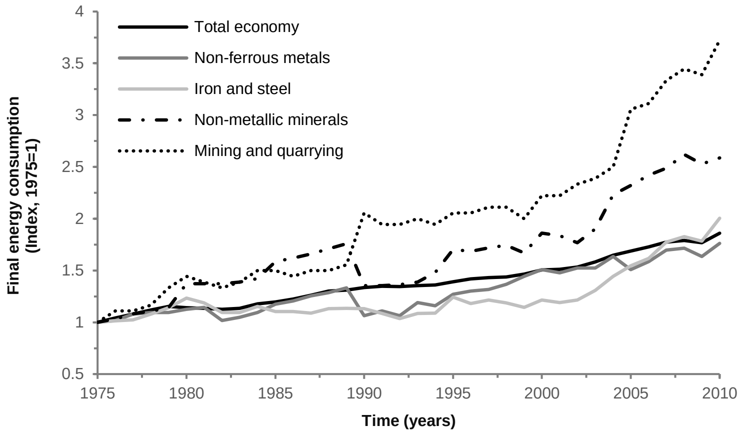


Figure 5.1 **Global energy consumption of metal sectors and total economy, 1975–2010. Data source: International Energy Agency (2015).**

### *5.1.3 Metal Ore Grade and Energy Cost of Extraction*

Indeed, economic rationality leads to first consuming metals from deposits where they are easily accessible and the least costly to extract (so most often where they are highly concentrated and close to the surface) and then to pursuing with less concentrated and deeper deposits. This implies that over time, the average ore grade of metal deposits decreases. This has been reported, both at a local deposit level (Crowson 2012), at national level (Mudd 2010; Mudd et al. 2013), and at worldwide level (Crowson 2012; Mudd et al. 2013). For instance, according to Mudd (2010), between the mid-nineteenth century and 2006, the average grade of copper in Australia fell from nearly 23% to less than 2%.

As theorized by Skinner (1976) the more you deplete a metal stock, the lower the concentration of the metal, the higher the unitary energy cost of extraction. More precisely, as the concentration of a given metal decreases, the energy cost associated with its extraction increases through an inverse mathematical relation of the power type. This relation has been precisely documented for copper, nickel and uranium (Mudd & Diesendorf 2008; Norgate & Jahanshahi 2010; Memary et al. 2012; Northey et al. 2013). In Figure 5.2, results of this relation from Norgate & Jahanshahi (2010) in the specific case of copper can be compared to a larger regression (used later in this chapter) operated on data presented in Table J.1 of Appendix J for 34 different metals.[1] The relation that is expressed in Figure 5.2 comes from an econometric regression, the results of which are presented in Table 5.1.

---

[1] The list of the 34 metals is as follows: Aluminum, Antimony, Arsenic, Beryllium, Bismuth, Cesium, Chrome, Cobalt, Copper, Gallium, Germanium, Gold, Hafnium, Indium, Iron, Lead, Lithium, Magnesium, Manganese, Mercury, Molybdenum, Nickel, Platinum, Rhenium, Silver, Tantalum, Tin, Titanium, Tungsten, Vanadium, Zinc, Zirconium, Praseodymium, and Neodymium.

**Table 5.1 Main results for the regression based on 34 metals represented in Figure 5.2.**

| Dependent Variable: Log(Energy consumption) | | | | |
|---|---|---|---|---|
| Method: Least Squares | | Observations included: 34 | | |
| Variable | Coefficient | Std. Error | t-Statistic | Prob. |
| C | 5.632090 | 0.240914 | 23.37803 | 0.0000 |
| Log(Ore Grade) | -0.600260 | 0.089179 | -6.730985 | 0.0000 |
| R-squared | 0.586061 | | | |
| Durbin-Watson | 2.167702 | | | |
| White test | 2.778711 | Prob. Chi-Square(2) | 0.2492 | |
| Jacque Berra | 2.767398 | Prob. | 0.2507 | |

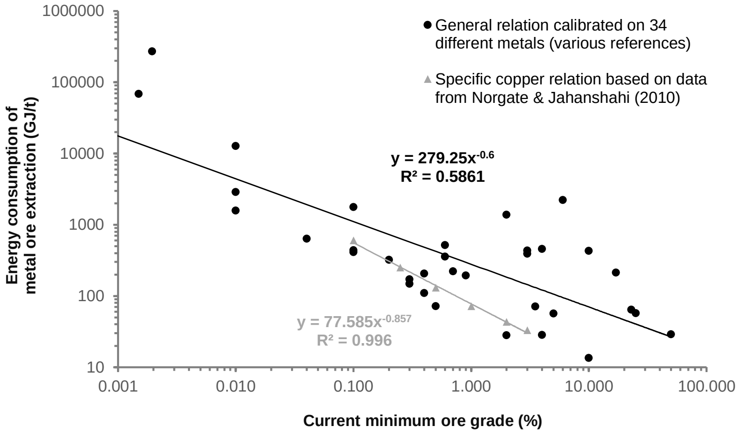


**Figure 5.2 Energy cost of metal extraction as a function of deposits' ore grades.**
**The grey line refers to copper only, the black line exhibits the same relation using an econometric regression on data for 34 different metals. Data source: Norgate & Jahanshahi (2010) for copper relation, Appendix J for general relation.**

The econometric regression shows that the relation between energy cost of extraction ($y$) and ore grade ($x$) can be estimated by the equation $y = ax^{-\alpha}$, with best estimate for $a = 279.25$ and $\alpha = -0.60026$ (and a 95% confidence interval of [-0.418609; -0.781910] for this same variable).

However, it must be stated that all metals will not follow the declining trend presented in Figure 5.2 at the same speed. Indeed, the speed of the ore grade degradation is different for each metal. Data from the USGS (2012) presented in Figure 5.3 show that the cumulative production of geochemically rare metals (such as gold, silver, copper, nickel, platinum, palladium, and rare earth elements) plus the reserves associated to these metals compared to their natural abundance in the Earth's crust is higher than for geochemically common metals (such as iron, aluminum, silicon, magnesium, manganese, and titanium). In this same figure, the natural abundance of the different metals (represented by the three thick lines) is obtained by multiplying the average grade of these metals in the continental crust by the mass of the continental crust in the top three kilometers, while the power fit regression (dashed lines) represents the relationship between the natural abundance of metals and their economic consumption (cumulative production plus reserves). Two points have to be mentioned here: first, the economic consumption of metals compared to their natural abundance is relatively imbalanced in favor of geochemically rare metals (comparison of the regression lines with the three thick lines). Second, between 1996 and 2012, the ratio of economic consumption to natural abundance increased faster for geochemically rare metals than for common metals (comparison of the slope of the two regression lines). It means, as already highlighted by Skinner (1976) that the depletion speed of rare metals accelerates more rapidly than for common metals.

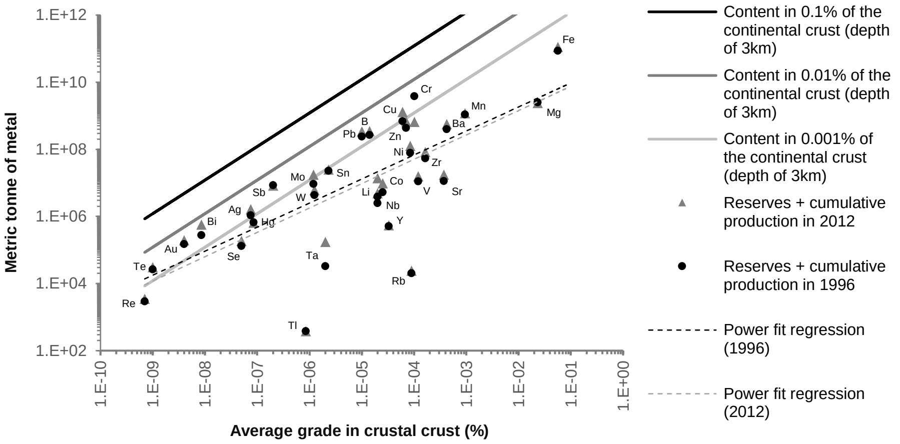


**Figure 5.3 Unbalanced anthropogenic consumption of metals, 1996–2012.**
**Reading: the graph shows that humans consume more rapidly geochemically rare rather than common metals. Indeed, the dashed regression lines represent the economic consumption of metals (cumulated consumption + reserves) while the thick lines show the natural abundance of the different metals in the continental crust. The difference between the black and grey dashed lines reveals the part of the natural abundance of metals which has been consumed for human needs between 1996 and 2012. Data source: USGS (2012).**

The question that then arises regards the impact of the energy cost of metal extraction on the energy balance of renewable technologies. Energy technologies are useful only if they are able to deliver more energy to society than is necessary to build and maintain their infrastructures and support their daily energy requirements, i.e. if their EROI remains above 1. Thus it is important, especially in the context of a complete transition towards renewable energy, to assess the sensitivity of the EROI to metal depletion.

## 5.2 EROI SENSITIVITY TO QUALITATIVE METAL DEPLETION

### *5.2.1 Equations and Data*

Two different calculations can be made in order to assess the impact of the energy cost associated with metal depletion on the EROI of a given electricity producing technology. A first approach consists in analyzing the individual contribution of a given metal’s energy cost of extraction on the EROI of this technology. Another related method explained thereafter can be used to determine the impact of a general quality exhaustion of all metals incorporated in the EROI of a given technology.

<u>Methodology for calculating the sensitivity of the EROI to one specific metal</u>

For a given electricity producing technology, it is possible to calculate the total energy production, $E_{out}$, from one MW of installed capacity during its entire lifetime $L$ by simply considering the load factor $\sigma$ of this technology and that there are 8760 hours in one year:

$$E_{out} = 8760 * \sigma * L. \tag{5.1}$$

Considering the current $EROI_{current}$, we can calculate the total energy invested, $E_{in}$, for one MW of this technology:

$$E_{in} = \frac{E_{out}}{EROI_{current}}. \tag{5.2}$$

And for each installed MW of the given technology, $I$ different metals have been extracted and consumed. As metal extraction does not account for the totality of the energy invested in the energy system, we define $\lambda_{current,i}$ as the ratio of the energy invested for the extraction of metal *i* and incorporated in the technology under study, $E_{\,current,i}$, over the total energy invested, $E_{in}$:

$$\lambda_{current,i} = \frac{E_{current,i}}{E_{in}}. \tag{5.3}$$

Similarly, $\lambda_{current,I}$, is the ratio of the energy invested for the extraction of all $I$ metals incorporated in the technology, over the total energy invested, $E_{in}$:

$$\lambda_{current,I} = \frac{\sum_i E_{\,current,i}}{E_{in}} \tag{5.4}$$

The current energy consumption due to the extraction of metal *i*, $E_{current,i}$, is obtained through real data by combining Table J.2 (metal intensity *i* of the technology in tons per MW, noted hereafter $\rho_i$) and Table J.1 (current unitary energy consumption of extraction of metal *i*, noted $\varepsilon_{current,i}$) of Appendix J. But the current unitary energy consumption, $\varepsilon_{current,i}$, can also be estimated thanks to its relationship to the current metal ore grade, $\tau_{current,i}$ as described in Figure 5.2 and here in (5.5):

$$\varepsilon_{current,i} = \mathrm{a}\tau_{current,i}{}^{-\alpha}. \tag{5.5}$$

Where $a$ and $\alpha$ are two coefficients that are estimated through econometric analysis in order for the relationship described in (5.5) to match real data as presented in Figure 5.2. Then, through this same relationship, we can compute the evolution of the unitary energy consumption of extraction of metal $i$, $\varepsilon_{evolved,i}$, if we suppose that the concentration of the ore grade of metal *i* has moved from $\tau_{current,i}$ to $\tau_{evolved,i}$:

$$\varepsilon_{evolved,i} = \mathrm{a} \times \tau_{evolved,i}{}^{-\alpha}. \tag{5.6}$$

Then, we can deduce the energy consumption due to the extraction of metal $i$ (from ore grade $\tau_{evolved,i}$) per MW of energy system installed, $E_{\,evolved,i}$ as a combination of the evolved unitary energy consumption previously calculated, $\varepsilon_{evolved,i}$, and the metal *i* intensity of the energy system, $\rho_i$:

$$E_{evolved,i} = \rho_i \times \varepsilon_{evolved,i}. \tag{5.7}$$

With (5.7), we can now compute the energy share, $\lambda_{evolved,i}$, of the metal *i* in $E_{in}$ under the assumption of an ore grade degradation of metal *i* from $\tau_{current,i}$ to $\tau_{evolved,i}$:

$$\lambda_{evolved,i} = \frac{E_{evolved,i}}{E_{in}}. \tag{5.8}$$

And we deduce $\lambda_{evolved,I}$, as the share of the energy invested in the electricity producing technology through the extraction of metal *i* with a degraded ore grade, and the extraction of all other metals except *i* (noted –*i*) operated at constant ore grade (i.e. current):

$$\lambda_{evolved,i,I} = \frac{\sum_{-i} E_{current,-i} + E_{evolved,i}}{E_{in}}. \tag{5.9}$$

Finally, we are able to calculate the $EROI_{evolved,i}$ of the technology under study, that is different from $EROI_{current}$ because of the ore grade degradation of metal *i* only:

$$EROI_{evolved,i} = \frac{EROI_{current}}{1 - \lambda_{current,I} + \lambda_{evolved,i,I}}. \tag{5.10}$$

By choosing different potential $\tau_{evolved,i}$ in a recursive process, we can calculate the sensitivity of the EROI of a given technology to any particular metal.

Methodology for calculating the general sensitivity of the EROI to all metals

If we want to calculate the sensitivity of the EROI of a given technology to all metals incorporated in such a technology, we have to make an assumption about the speed of exhaustion of the different metals. Indeed, as previously stated the speed of this evolution will differ from one metal to another and depends mostly on the *Clarke Value* of the metal considered (Craig et al. 2001; Valero & Botero 2002; Rankin 2011). This indicator is defined at a given time as the ratio of the minimal concentration of a given metal that is economically acceptable for exploitation (in the current period of exploitation) to its average concentration in crustal crust. Figure 5.4 exposes the relation established between the Clarke Value of copper and the multiplying factor that would affect the energy cost of copper extraction if the average concentration of this metal were to go from its economic minimal concentration to its average crustal crust concentration (based on data from Norgate & Jahanshahi, 2010). This analysis is extended to the 34 different metals previously used (Figure 5.4).

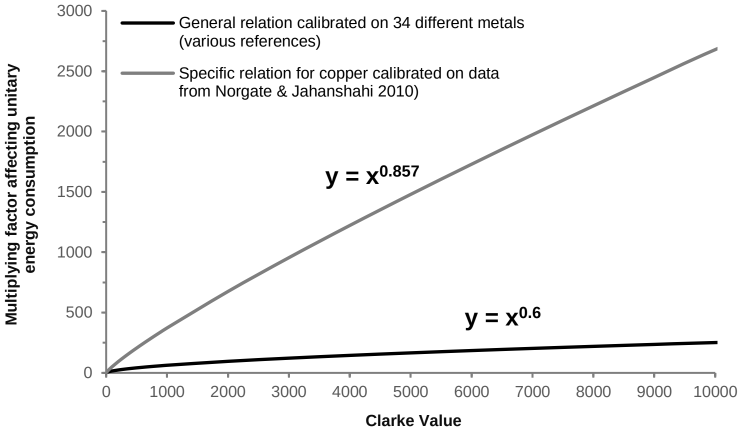


**Figure 5.4 Increasing energy cost of metal extraction as a function of Clarke Values. The grey curve represents the relation for copper only, the black curve represents the same relation calibrated on 34 different metals. Data source: Norgate & Jahanshahi (2010) for copper relation, Appendix J for general relation.**

Figure 5.4 can be read as follows with the example of copper (grey line): a typical copper deposit of minimal profitability has an ore grade of 0.5%, whereas copper grade in common rocks is about 0.006%. As a consequence, the Clarke Value of copper is approximately 83 (0.5/0.006=83.33). Thus, exploiting copper from common rocks instead of concentrated deposits would imply multiplying the current energy cost of extraction by 45 (83^0.857=44.12). By a way of comparison in financial terms, Steen & Borg (2002) have shown that if metals were extracted from common rocks, the financial cost associated with such an exploitation would be multiplied by a factor 10 to 10,000.

Thus, in order to build a methodology to test the sensitivity of the EROI to all metals, we make the convenient assumption that all metals that are considered are depleted in the same proportion. As a consequence we only consider geochemically rare metals because geochemically common metals have low Clarke Values, implying that for them a shift from concentrated deposits to common rocks would not induce a great change in their energy cost of extraction. Thus, in the following section we only consider the subset $J$ of geochemically rare metals among the $I$ different (rare and common) metals incorporated in the technology under study. Furthermore, we suppose that their ore grade is equally divided by a factor $\theta$ through a given period of time of the extraction process. In this context, the relationship linking the current ore grade of rare metal $j$, $\tau_{current,j}$, to its initial unitary energy consumption, $\varepsilon_{current,j}$, is provided by:

$$\varepsilon_{current,j} = \mathrm{a} \times \tau_{current,j}{}^{-\alpha}. \tag{5.11}$$

We wish to understand how the energy consumption associated with the extraction of the rare metal $j$ is modified and equals $\varepsilon_{evolved,j}$ when its ore grade is reduced by a factor $\theta$:

$$\varepsilon_{evolved,j} = a \times (\tau_{current,j}/\theta)^{-\alpha}. \tag{5.12}$$

We divide (5.12) by (5.5) and obtain the multiplying factor $\mu_j$ affecting the energy consumption per unit of extracted metal $j$ when its ore grade is divided by the factor $\theta$:

$$\mu_j = \frac{\varepsilon_{evolved,j}}{\varepsilon_{current,j}} = \theta^{\alpha}. \tag{5.13}$$

Then, we combine (5.13) with the previous equation (5.4) in order to obtain (5.14), where $\lambda_{evolved,J}$ represents the share of the energy required for the production of all the different $J$ geochemically rare metals over the total energy invested in the energy system once all (rare) metals ore grades have been diminished by a factor $\theta$:

$$\lambda_{evolved,J} = \frac{\sum_j \mu_j \times E_{current,metal\ j}}{E_{in}}. \tag{5.14}$$

We are now able to calculate the evolution of the EROI, now called $EROI_{evolved,J}$, of the energy technology due to the degradation of all $J$ rare metals concentration by the same factor $\theta$:

$$EROI_{evolved,J} = \frac{EROI_{current}}{1 - \lambda_{current,J} + \lambda_{evolved,J}}. \tag{5.15}$$

By choosing different factors $\theta$ in a recursive process, we can calculate the sensitivity of the EROI of a given technology to all geochemically rare metals.

Data requirement for numerical applications

If we wish to perform numerical applications, both methodologies previously presented require data concerning: energy cost of metal extraction, metal requirement per electricity producing technologies, EROI, load factor and capital lifetime. Examples of such assumptions are proposed in Table 5.2 (EROI, load factor and capital lifetime), Table J.1 (energy cost of metal extraction), and Table J.2 (metal requirement per electricity producing technology) of Appendix J.

**Table 5.2 Current EROI, load factor, and lifetime of electricity producing technologies.**

| **Technology** | **Load factor, $\sigma$ (%)** | **Lifetime, $L$ (years)** | **Current EROI (X :1)** | **Reference for EROI** |
|---|---|---|---|---|
| Parabolic through | 33 | 30 | 20 | Weißbach et al. (2013) |
| Solar tower plant | 33 | 30 | 20 | Kreith & Krumdieck (2014) |
| PV single Si | 10 | 25 | 6 | Raugei et al. (2012) |
| PV multi Si | 10 | 25 | 6 | Raugei et al. (2012) |
| PV a Si | 10 | 25 | 4 | Raugei et al. (2012), Weißbach et al. (2013) |
| PV CIGS | 10 | 25 | 6 | Raugei et al. (2012), Weißbach et al. (2013) |
| PV CdTe | 10 | 25 | 12 | Raugei et al. (2012) |
| Onshore Wind | 25 | 20 | 18 | Kubiszewski et al. (2010) |
| Offshore Wind | 35 | 20 | 18 | Kubiszewski et al. (2010) |
| Hydropower | 60 | 100 | 50 | Weißbach et al. (2013) |
| Nuclear | 80 | 40 | 10 | Hall & Day (2009) |

### *5.2.2 Results of Simulations*

Impact of specific metal scarcity on the EROI of different electricity producing technologies

The methodology developed above allows one to calculate the impact of the degradation of a specific metal ore grade on the EROI of different electricity producing technologies. Such a calculation is in principle feasible for any metal that is used in a given technology but for the sake of clarity, the results concerning only three metals are presented: copper (Figure 5.5), nickel (Figure 5.6), and chromium (Figure 5.7).

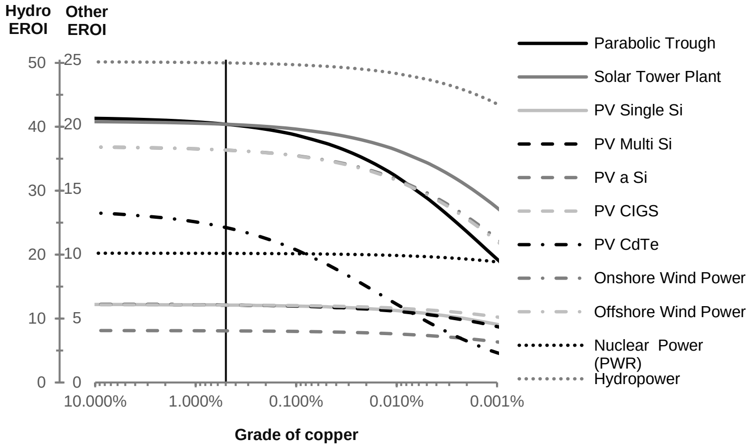


**Figure 5.5 Sensitivity of the EROI of different energy technologies to the grade of copper. Relationship: energy cost of metal extraction=1.397*grade^-0.60026. The vertical black line represents current grade.**

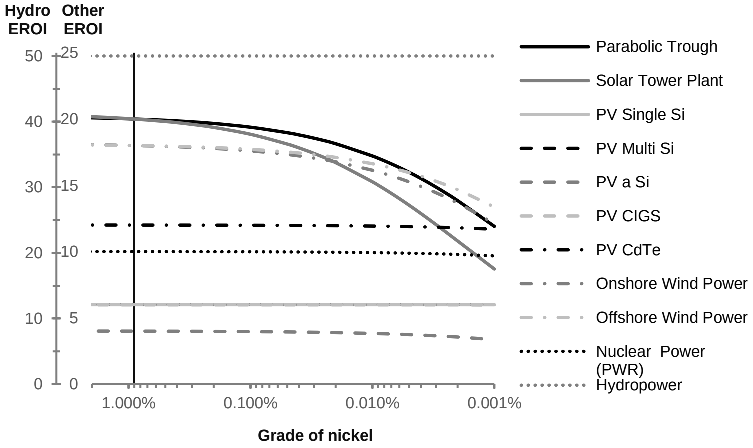


**Figure 5.6 Sensitivity of the EROI of different energy technologies to the grade of nickel. Relationship: energy cost of metal extraction =11.463*grade^-0.60026. The vertical black line represents current grade.**

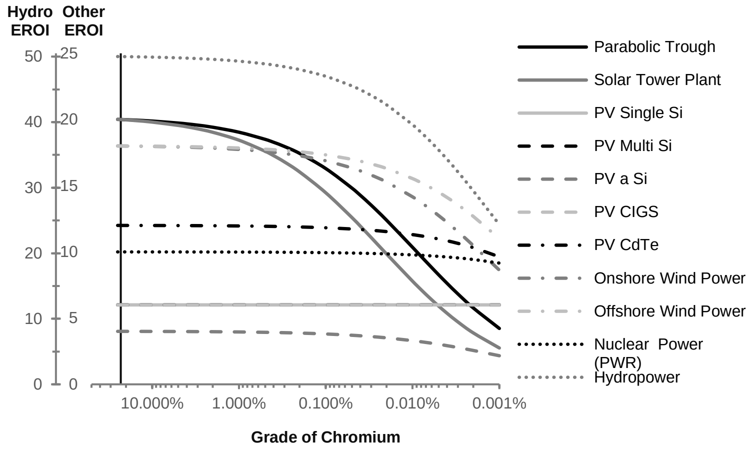


**Figure 5.7 Sensitivity of the EROI of different technologies to the grade of chromium. Relationship: energy cost of metal extraction =26.529*grade^-0.60026. The vertical black line represents current grade.**

A comparison of these different figures indicates that technologies are not equally sensitive to the three different metals (copper, nickel, and chromium) chosen as examples. For instance, hydropower is more sensitive to chromium than copper because at the same extremely low concentration of 0.001%, the EROI of hydropower is lower in the case of chromium ore grade degradation than in the similar case for copper. But one could say that because chromium is currently exploited in deposits with high concentrations (23%) compared to copper (0.5%), the EROI of the different technologies will probably be impacted first by copper rather than chromium grade degradation. Trying to say that the EROI of a technology is more sensitive to a given metal compared to another depends not only on the level of diminution of its EROI, but also on the time at which this impact will start. This time horizon problem is out of scope here as it would require building complex scenarios relying on different assumptions (GDP and population growths, intensity in the use of the different metals in the energy system and in other societal uses, etc.). By way of illustration, Crowson (2012) has provided some data about the evolution of the grade of copper. According to this author, in 1800, the economical copper mines of the United of Kingdom were characterized by an average copper grade of nearly 9.27% and as previously stated an average value of 0.5% is now characteristic of copper mines.

Only one peer-reviewed study from Harmsen et al. (2013) has investigated the relation between energy cost of metal extraction and EROI. In their analysis, Harmsen et al. have investigated how the evolution of copper consumption and its associated energy cost of extraction could affect the EROI of wind turbines on a 2050 horizon, assuming different 100% renewable energy scenarios. Their results showed that the EROI of wind turbines would be marginally impacted (3% of the original value) by copper consumption on this time period if only wind turbine systems are studied. Taking into account grid and backup needs would more importantly impact the EROI of the energy system (15% decrease compared to initial value).

### Impact of general metal scarcity on the EROI of different electricity producing technologies

An alternative to the first methodology was developed in order to calculate the sensitivity of the EROI of the same technologies to all metals, considering a common degradation of their deposit's concentration. The results of these calculations are presented in Figure 5.8.

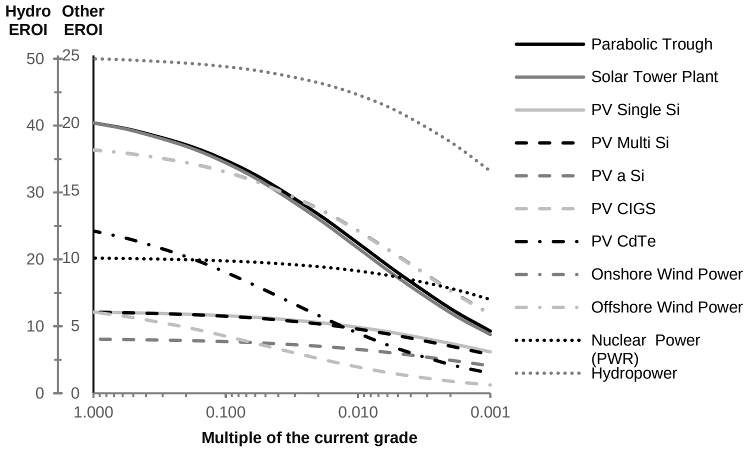


**Figure 5.8 EROI sensitivity to general qualitative depletion ($\alpha = 0.60$).**
**Reading: a multiple of the current grade of 0.1 means that the current grades of all geochemically rare metals are divided by a factor of 10. Relationship $\mu = \theta^{\alpha}$ where $\alpha = 0.60026$.**

In this case, all technologies are affected but not equally, which is rather logical. Some differences are worth pointing out. In Figures 5.5, 5.6, and 5.7, the evolution of the EROI of PV Multi Si and PV CIGS is pretty much the same and differences in impacts are hardly discernible. On the other hand, the evolution of the EROI of onshore and offshore wind power shows discrepancies in Figure 5.7 (sensitivity to chromium), whereas it exhibits the same behavior in Figure 5.8 when all rare metals are accounted for, highlighting the existence of compensatory effects. This shows that taking into account all metals is important for a deep understanding of the impact of metal scarcity on the EROI of energy systems.

These results also show that if rare metals were extracted from deposits with ore grade approaching very low concentration (as we move to the far right of Figure 5.8), the energy requirement would be so important that it would considerably decrease the EROI of all electricity producing technologies. Under such a scenario, only a few renewable technologies (hydro and wind power) and nuclear would still present EROI well above the breakeven point. In such a situation, wind turbines would still deliver net energy to society but as shown in Section 4.3.2, this would surely imply a large increase in electricity prices.

### *5.2.3 Sensitivity Analysis*

As exposed before, in order to assess the sensitivity of the EROI of energy technologies to metal grade depletion, the general econometric relation presented in Figure 5.2 was used. As a consequence, the results previously presented are particularly sensitive to the value of parameter $\alpha$. So far, results have been presented using the best estimate for this parameter (-0.60026). In particular, lower ore grades are underestimated with this mean $\alpha$, whereas using the upper estimate of the 95% confidence interval (that is 0.781910 instead of 0.60026) would put more weight on lower grades. Figure 5.9 presents the same results as in Figure 5.8 regarding the sensitivity of the EROI to the ore grade degradation of all rare metals but with an $\alpha$ of 0.781910. Under such conditions, the EROI of all technologies are logically more sensitive to important ore grade degradations. In Figure 5.9, the decrease in all EROIs is greater and occurs at lower ore grades.

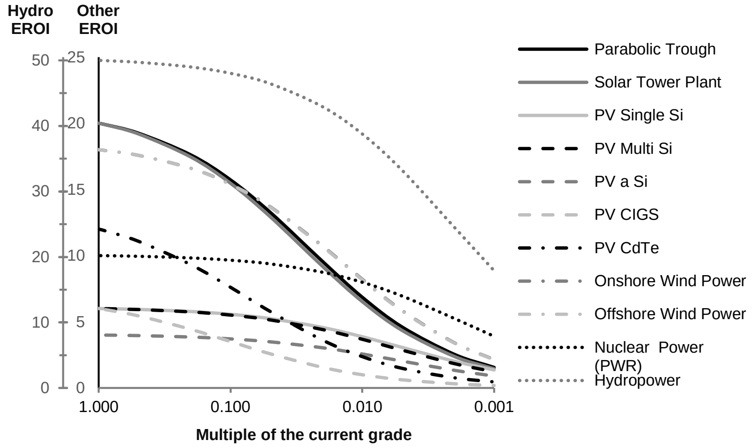


**Figure 5.9 EROI sensitivity to general qualitative depletion ($\boldsymbol{\alpha = 0.78}$).**
**Reading: a multiple of the current grade of 0.1 means that the current grades of all geochemically rare metals are divided by a factor of 10. Relationship: $\boldsymbol{\mu = \theta^{\alpha}}$ where $\boldsymbol{\alpha = 0.781910}$.**

## 5.3 QUALITATIVE METAL DEPLETION IN THE ENERGY TRANSITION CONTEXT

### *5.3.1 Potential Vicious Circle Arising Between Energy and Metal Sectors*

As highlighted in Section 4.2.1, it appears that energy resources to which modern economies have become accustomed and on which they are dependent do not generate as much net energy as they used to do. Indeed, all fossil fuels present a declining EROI trend and unconventional fossil fuels present relatively low EROI. Renewable technologies with which policy makers would like to replace these stock-based energy resources present EROIs that are (currently) lower than past fossil fuel EROIs. There is still room for technological progress to increase the EROI of these renewable technologies, but it will ultimately encounter a limit; and as already stated, renewable technologies are more capital-intensive, and in particular metal-intensive, than conventional means of energy production. As depletion occurs for metals, the energy cost associated with their extraction increases following a highly non-linear pattern (inverse power function). In the context of a transition toward renewables, all other things being equal, the increasing energy requirement of the metal sector due to metal ore grade degradation would further increase the demand for renewable technologies. Moreover, the intermittency of these technologies implies the need to expand and reinforce the transmission grid, thus generating an even greater demand for metals. As a consequence, in the perspective of a transition toward renewable technologies, a potential vicious circle could develop between energy and metal sectors as summarized in Figure 5.10.

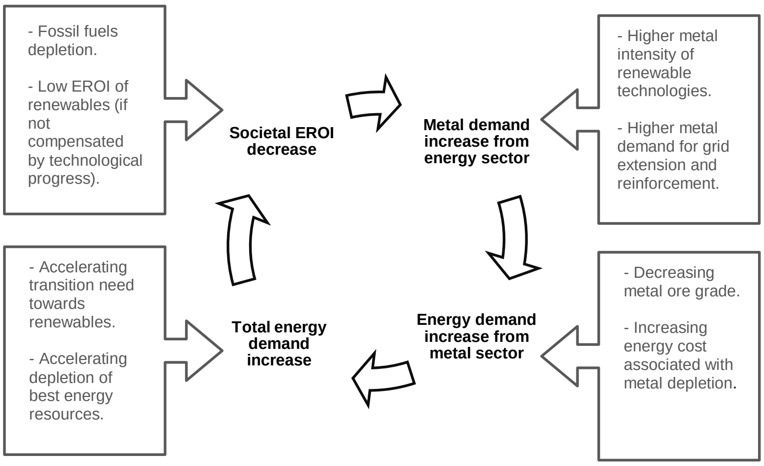


**Figure 5.10 Potential energy-metal vicious circle of the coming energy transition.**

### *5.3.2 Attenuating Levers of the Self-Enhancing Relation between Energy and Metal Sectors*

This self-enhancing relation between energy and metal sectors depicted in Figure 5.10 could be attenuated thanks to different levers. Recycling could slow down metal depletion but also decrease the energy requirement of the metal sector. Norgate & Haque (2010) gave the estimate of potentially 65 to 95% energy savings. However it must be stated that the effect of recycling is limited when the economy experiences continuous high growth. Moreover, a 100% recycling efficiency is impossible due to physical dissipation as supported by the second law of thermodynamics (Ayres 1999; Craig 2001). A dematerialization of the energy sector implying the consumption of less matter per MW of installed capacity is another possible channel of improvement. However, it must be stated that final energy will always needs a minimum physical-based support to exist. Gains in energy efficiency in order to lower the energy requirement for metal extraction are also a solution. According to different commentators (Ruth 1995; van der Voet et al. 2013; Norgate & Jahanshahi 2010; 2011) such sources of energy efficiency are quite large in the metal sector, though ultimately limited by thermodynamic laws. In that regard, Ayres (2007) estimates that technological progress could reduce the energy required for the extraction and concentration of copper by no more than a factor of two or three.

Technical substitution of rare metals with common metals (Skinner, 1976) and others materials (Goeller & Weinberg 1978) is another lever. Nonetheless, as demonstrated by Messner (2002), incentives for substitution triggered by the price signal do not always lead to a modification of the production technology due to the presence of large switching costs. Moreover, the substitution process is affected by inertia and the research into substitutes (perfect or imperfect) needs time, money, and adequate economic incentives. Energy economies of scope through metal coproduction are another option. As economies move on to less concentrated deposits, the opportunity of exploiting deposits with multiples metals coproduction will appear more advantageous than nowadays thanks to the possibility of scope economies. Indeed, metals produced as byproducts or coproducts (mainly minor metals) benefit from an energy credit as the energy cost associated with extraction and concentration is completely allocated to the primary ore and not to the byproduct. As a consequence, only the refining energy cost is allocated to minor metals. Nonetheless, even in the case of coproduction with only one primary metal that supports the energy cost of concentration, the exhaustion causes an important effect on the EROI (see graphs on individual sensitivity). In the same way, economies of scale can be a good way to reduce the unitary energy consumption even if the general consequences of this kind of measure are largely unknown in terms of ore grade degradation and total mining energy consumption.

### *5.3.3 Enhancing Factors of the Self-Enhancing Relation between Energy and Metal Sectors*

On the other hand, other factors could accelerate and enhance the relation presented in Figure 5.10. Skinner (1976) enunciated what he called a "mineralogical barrier". Under a certain threshold metal would not appear as "grains" in mineral but would substitute other atoms in the crystalline structure.[1] In this case, you would have to chemically break the totality of the mineral to recover the desired metal, which would prove to be really expensive from an energy perspective. Skinner (1976) even evoked a break in the relation between ore grade and energy cost of extraction (also discussed in Ayres, 2007). Future deposits that will be put into production will be deeper, and will probably contain more impurities. This will require more energy to convey ore to the surface and to operate a finer crushing (van der Voet et al. 2013). Other environmental impacts have not been considered so far, such as waste management, water needs, or greenhouse gases emissions, etc. If the management of such negative externalities were to be integrated, it would surely imply an additional energy cost. As already expressed by Harmsen et al. (2013), the energy cost associated with the construction and maintenance of other parts of the energy system, related to the transmission and distribution grid or to electricity storage, would induce a further reinforcement of the relation between energy and metal sectors previously depicted.

This fifth chapter shed some light on the close relationship between the energy and metal sectors from the EROI perspective. First, we have supported the position of Rankin (2011) by estimating that 10% of global primary energy production is consumed by the metal sector. Then, we showed that the energy consumption of the metal sector has increased faster than the rest of the economy since 1973. Supported by previous studies, we made the fair assumption that this apparent increasing energy requirement of the metal extraction sector is mainly due to decreasing ore grade. The decline in quality of ores is a natural process that occurs across different scales (deposit, nation, and world) and implies that more energy is needed to extract a given quantity of metal. Because renewable technologies have higher metal intensities than conventional means of electricity production, the question of the sustainability of a transition consisting in a shift toward renewables is legitimate, especially because those energy systems have lower EROIs than fossil fuels. Logically, we have decided to estimate how the energy requirement associated with metal extraction could impact the EROIs of different electricity producing technologies.

---

[1] For example, lead is a substitute for potassium at atomic scale, as is zinc for magnesium.

A first analysis has consisted in calculating the sensitivity of the EROI of renewable and nuclear technologies assuming different levels of ore grade degradation for a specific metal. We have explained the kind of results that it is possible to obtain through the example of three metals (copper, nickel, and chromium), although this kind of sensitivity calculation could have been performed for any metal used in a given technology. Each technology displays a specific sensitivity to a particular metal that can be measured through the methodology we have developed. In a second step, we have adapted our methodology in order to calculate the sensitivity of the EROI of the same technologies to a similar depletion of all rare metals. This exercise was useful to see that energy requirements associated with metal extraction could have a significant impact on the capacity of these “green” technologies to deliver net energy to society. Of course, the question of the speed of degradation of the average ore grade of a given metal remains unanswered. This evolution will be different for each metal but will ultimately have a negative impact on the EROI of renewable technologies.

In the context of a transition toward renewables that are more metal-intensive than fossil energy systems, all other things being equal, the increasing energy requirement of the metal sector due to metal ore grade degradation will further increase the demand for renewable energy. Moreover, the intermittency of these technologies implies the need to expand and reinforce the transmission grid and storage capacities, which will generate an even greater demand for metals. As a consequence, in the perspective of a transition toward renewable technologies, a potential vicious circle could developed between the energy and metal sectors. It is currently impossible to say whether such an unpleasant situation would effectively arise but this chapter has started a quantitative exploration of this issue.

# CHAPTER 6

# ENERGY EXPENDITURE, EROI, AND ECONOMIC GROWTH

*"The farther back you can look, the farther forward you are likely to see."*

Winston Churchill

As shown in the first three chapters of this thesis, there is no consensus on the relative contributions of production factors to economic growth. The attention paid by classical economists to land vanished when modern industrial growth shifted the emphasis to capital availability, whereas the importance of routine labor and human capital has never been questioned, probably simply because economics is by essence the study of a human system in which humans *must* play the leading part. But in Chapter 3 we saw that the fundamental laws of thermodynamics must necessarily apply to the economic system so that the role of energy is indeed crucial to the growth process. In Chapter 4 we investigated in more details how energy has been taken into account in the biophysical paradigm with the related concepts of net energy and EROI; and in Chapter 5 we saw how the qualitative depletion of metals could possibly affect the future evolution of energy system's EROI. In the present chapter, we will first see that the econometric literature investigates the relation between energy and economic growth from two possible perspectives: the relation between energy *price* and economic growth, or the relation between energy *quantity* and economic growth. Then, we will see that these approaches can be combined with the notion of *energy expenditure*, generally expressed as a share (or fraction) of GDP. Energy expenditures level are then computed for the US and world economy from 1850 to 2012, and for the UK from 1300 to 2008. In a third step we give, for the US only due to data availability and consistency, an estimation of the ultimate level of total energy expenditure (as a fraction of GDP) above which economic growth seems statistically impossible. Then, this result is expressed as (i) the maximum tolerable aggregated energy price (and oil price), and (ii) the minimum required EROI that the energy sector must have in order for the US economic growth to be positive. Finally, we give the results of various Granger causality tests performed on the US case for the restricted period 1960–2010.

## 6.1 CONTRIBUTION OF ECONOMETRICS TO THE ENERGY-ECONOMIC GROWTH NEXUS

### *6.1.1 Energy Price and Economic Growth*

Hamilton (1983) was the first of a score of studies concentrating on the relation between energy prices (usually the oil price) and economic growth (Lardic & Mignon 2008; Katircioglu et al. 2015). Because the oil price impacts economic growth asymmetrically,[1] the classical methods of cointegration are ineffective, and more sophisticated methods are required to evaluate the energy price–economic growth relation (Lardic & Mignon 2008; An et al. 2014). The scarcity of data on energy prices (across different countries and over time) complicates the assessment of this relation. In a nutshell, this literature seems to converge toward a feedback relation between variations in energy price and economic growth (Hanabusa 2009; Jamil & Ahmad 2010), ranging from a negative to a positive effect depending on the level of oil-dependency of the country under study (Katircioglu et al. 2015); and a clear negative inelastic impact of the oil price on GDP growth rates for net oil-importing countries. In addition, Naccache (2010) has shown that the impact of the energy price on economic growth depends on the origin of the oil price shock (supply, demand, or pure speculative shock), taking into account that the relative importance of each of these shock-drivers has varied considerably over time (Benhmad 2013). When reviewing the literature, we found that all these studies consider that the oil price can exert a constant effect on an economy between two dates, whereas the energy intensity of this economy may obviously vary greatly over the same period of time. Just as the studies rightly assume that low- and high-energy intensive countries would not react in exactly the same way when confronted with increased energy prices, (because the former are clearly less vulnerable), the same point should also be taken into account for a given country studied at different times. We therefore recommend explicitly introducing energy intensity as a key variable in future diachronic empirical assessments of energy price–economic growth relations.

### *6.1.2 Energy Quantity and Economic Growth*

Another impressive array of studies focuses on the relation between quantities of energy consumed and economic growth. Such studies have been conducted since the seminal paper of Kraft & Kraft (1978). From this energy quantity–economic growth nexus, four assumptions have been envisaged and systematically tested:

- A relation of cause-and-effect running from energy to economic growth. Studies supporting this assumption come close to the thinking of the biophysical movement

---

[1] The asymmetric response of the economy to the variation of the oil price can be explained by different factors such as the monetary policy, the existence of adjustment costs, the presence of uncertainty affecting investment choices and the asymmetric response of oil-based products to oil price variations. In the case of an oil price variation, the different adjustment costs may result from sector shifts, change in capital stock, coordination problems between firms, and uncertainty. When combined, these adjustment costs can completely erase the benefits associated with a fall in the oil price. See Lardic & Mignon (2008) and also Naccache (2010) for more information.

(presented in the following subsection) and the proponents of peak oil, because it gives a central role to energy in the economic process.

- A causal relation running from economic growth to energy. In this situation, energy is not essential and energy conservation policies can be pursued without fear of harming economic growth. This conservative view reflects the position of many neoclassical economists for whom energy is seen as a minor and easily substitutable production factor.
- A feedback hypothesis between energy and economic growth.
- The absence of any causal relation between energy and economic growth, which is also known as the neutrality assumption.

After more than forty years of research and despite the increasing sophistication of econometric studies, this area of study has not so far led to either general methodological agreement or a preference for any of the four positions. More specifically, three independent literature reviews (Chen et al. 2012; Omri 2014; Kalimeris et al. 2014), covering respectively 39, 48, and 158 studies, have shown that no particular consensus has emerged from this empirical literature and that the share of each assumption ranges from 20% to 30% of the total. Various explanations can be suggested for these mixed results, including the period under study, the countries in question (the level of development affecting the results), the level of disaggregation of the data (GDP or sectorial levels), the type of energy investigated (total energy, oil, renewable, nuclear, primary vs. final energy, exergy, etc.), the econometric method applied (OLS, cointegration framework, VAR, VECM, time series, panel or cross-sectional analysis), the type of causality tests (Granger, Sims, Toda and Yamamoto, or Pedroni tests), and the number of variables included in the model (uni-, bi-, or multivariate model) (Huang et al. 2008a; 2008b; Kocaaslan 2013; Fondja Wandji 2013).

### *6.1.3 Energy Expenditure and Economic Growth*

The two oil shocks of the 1970s were stark reminders of the world economy's dependence on fossil energy. Energy expenditure, also called energy cost, is the quantity of economic output that must be allocated to obtaining energy. It is usually expressed as a fraction of Gross Domestic Product (GDP). Murphy & Hall (2011a; 2011b) suggest that "when energy prices increase, expenditures are re-allocated from areas that had previously added to GDP, mainly discretionary investment and consumption, towards simply paying for more expensive energy". These authors show graphically that, between 1970 and 2007, the economy of the United States of America (US) went into recession whenever the petroleum expenditure of the US economy exceeded 5.5% of its GDP. In addition, Lambert et al. (2014) suggest that in the US once energy expenditure rises above 10% of GDP recessions follow.

Bashmakov (2007) makes a difference between energy cost to GDP ratio and energy cost to final consumer income ratio. He identifies energy cost to GDP thresholds of 8–10% for the US (4–5% for final consumer income) and 9–11% for the OECD (4.5–5.5% for final consumer income) below which he finds almost no correlation between the burden of energy expenditure and GDP growth rates. However, when these thresholds are exceeded, the economy slows down and demand for energy falls until the energy cost to GDP/consumer income ratios are back below their thresholds. Bashmakov (2007) argues that until the ratio of energy expenditure to GDP reaches its upper critical threshold, it is all the other production factors that determine the rates of economic growth, and energy does not perform a "limit to growth" function. "But when energy costs to GDP ratio goes beyond the threshold, it eliminates the impact of factors contributing to the economic growth and slows it down, so the potential economic growth is not realized".

King et al. (2015b) estimate energy expenditures as a fraction of GDP for the period 1978–2010 for 44 countries representing 93–95% of the gross world product (GWP) and 73–79% of the IEA's listed world Total Primary Energy Supply (TPES) (>78% after 1994). The methodology used by these authors is set out in full in their article but it should be pointed out that they consider coal, oil, and natural gas for three sectors (industrial, residential, and electricity production), plus non-fossil (nuclear, renewable) electricity production for two sectors (industrial and residential). The quantities and prices of these different commodities were mostly retrieved from databases of the US Energy Information Administration (EIA). King et al. (2015b) aggregate these national energy costs to estimate the global level of energy expenditure from 1978 to 2010. They find that this estimated energy cost as a fraction of the GWP fell from a maximum of 10.3% in 1979 to 3.0% in 1998 before rising to 8.1% in 2008. King (2015) uses these data to perform simple econometric correlation (hence not causal) analyses that deliver the following main results: expenditure on energy expressed as a fraction of GDP is significantly negatively correlated with the one-year lag of the annual changes in both GDP and total factor productivity, but not with the zero-year lag of these same variables.

As already stressed, the various energy expenditures estimated by King et al. (2015b) were only for the period 1978–2010, and the econometric analyses of King (2015) were not designed to infer any temporal causality between energy expenditure and economic growth, nor to estimate any potential threshold effect in such a relation. Consequently, we seek to achieve two related goals in the present chapter. First, we think it is important to extend energy expenditure estimates (as fractions of GDP) to a larger time frame, for as many countries as possible.[1] In the following section we are able to do this adequately for the US and the global economy from 1850 to 2012, and for the United Kingdom (UK) from 1300 to 2008.[2] Second, we wish to relate the level of energy expenditure as a fraction of GDP to the economic growth dynamics in order to quantitatively support the various qualitative results previously advanced by Murphy & Hall (2011ab), Lambert et al. (2014), and King (2015). More precisely, focusing on the US due to the availability and consistency of data, we seek to:

- (i) Estimate the ultimate level of energy expenditure (as a fraction of GDP) above which economic growth statistically vanishes.

---

[1] Fouquet (2011) highlights the danger of focusing on the price of energy rather than the price of energy services when considering the long-run because the former ignore major technological improvements. We completely agree with this statement and want to highlight that our work takes into account some of this technological progress through the energy intensity of the economy.

[2] Naturally, the geographical definition of the United Kingdom is quite blurred over such long time span (see Fouquet 2008 for details).

- (ii) Express this result in terms of the maximum average price of energy and the minimum societal energy-return-on-investment (EROI) that must prevail in the economy in order for economic growth to be positive.
- (iii) Perform Granger causality tests to identify the direction of the possible causal relation between energy expenditure and GDP growth.

## 6.2 ENERGY EXPENDITURE IN THE VERY LONG RUN

### *6.2.1 Equations and Boundary to Estimate Energy Expenditures*

We note $X_j$ the level of expenditure of a given energy *j* produced in quantity $E_j$ and sold at price $P_j$ in a given economy:

$$X_j = P_j E_j. \tag{6.1}$$

In our study, the *j* energy forms include the following marketed energy: coal, crude oil, natural gas, non-fossil electricity (i.e. nuclear and renewable electricity from hydro, wind, solar, geothermal, biomass and wastes, wave and tidal) and modern biofuels (ethanol and biodiesel). Hence, total expenditure of marketed energy, $X_{total\ marketed}$, is:

$$X_{total\ marketed} = \sum_j X_j = P_{average} E_{total\ marketed}. \tag{6.2}$$

With $P_{average}$ as the quantity-weighted average price of aggregated marketed energy:

$$P_{average} = \sum_j P_j \frac{E_j}{\sum_j E_j}, \tag{6.3}$$

and $E_{total\ marketed}$ the total supply of marketed energy:

$$E_{total\ marketed} = \sum_j E_j. \tag{6.4}$$

Usually, such estimates of marketed energy expenditure omit traditional biomass energy (woodfuel, crop residues[1]) because they usually represent non-marketed consumption for which average annual prices cannot be estimated. Consequently, if such an energy resource is omitted from equations (6.1) and (6.2), we necessarily underestimate contemporary levels of energy expenditure since woodfuel and crop residues still represent 70% of global renewable energy consumption nowadays (whereas hydro accounts for 20% and new renewable technologies such as wind power, solar PV, geothermal, wave, tidal, wastes, and modern biofuels account for the remaining 10%). But most importantly, for times prior to the 1940s when traditional biomass represented a large share of the total primary energy supply of

---

[1] Formally, fodder supplied to draft animals should be added to traditional biomass energy estimates, but it is generally discarded due to difficulties of estimation. This is also the case for traditional windmills and water wheels.

many countries, we need a proxy for total energy expenditure including non-marketed energies in order to have a more accurate idea of the actual level of total energy expenditure. With $E_{trad}$ as the quantity of traditional biomass energy, and $TPES = E_{total\,marketed} + E_{trad}$ as the total primary energy supply, we define, for a given economy, the proxy of total energy expenditure, $X_{total\,proxy}$, as:

$$X_{total\,proxy} = \frac{X_{total\,marketed}}{\left(1-\frac{E_{trad}}{TPES}\right)}. \tag{6.5}$$

In our results we will present a (second best) estimate of total energy expenditure for the US and world economy using the total proxy method in order to test its consistency with the (first best) estimate which includes woodfuel as marketed energy.

### *6.2.2 Data for the USA, the UK, and the World*

Data for the US

We used several sources summarized in Table 6.1 in order to estimate the prices of coal, crude oil, gas, electricity, woodfuel, and modern biofuels consumed in the US. In order to express all energy prices in the same convenient unit, i.e. 1990 International Geary-Khamis dollars[1] per terajoule (abbreviated \$1990/TJ, where 1 TJ = $10^{12}$ J), we used the US Consumer Price Index of Officer & Williamson (2016) and different energy conversion factors from British Petroleum (2015) such as the average energy content of one barrel of crude oil (5.73E-03 TJ), the average energy content of one metric tonne of hard coal (29.5E-03 TJ), the average energy content of one thousand cubic feet of natural gas (1.05E-03 TJ), the average energy content of one gasoline gallon equivalent (1.2E-04 TJ), the average energy content of one thousand board feet of wood (2.3E-02 TJ), and the terajoule equivalent of one kWh (3.6E-06). We present in Figure 6.1 the resulting prices of coal, oil, gas, and woodfuel expressed in \$1990/TJ, and electricity in thousand \$1990/TJ (biofuels prices are omitted from this figure for the sake of clarity).

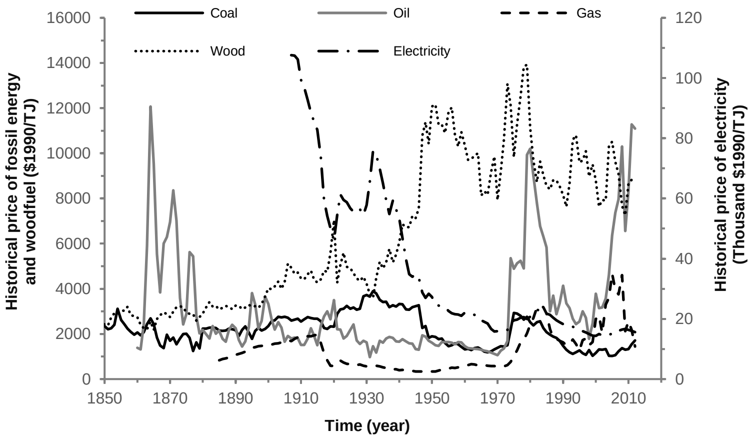


**Figure 6.1 Estimations of US energy prices in \$1990/TJ, 1850–2012.**
**Left scale in \$1990/TJ for coal (1850–2012), oil (1860–2012), gas (1890–2012), and woodfuel (1850–2012); and right scale in thousand \$1990/TJ for electricity (1907–2012).**

---

[1] For definition see Section 1.1.1, footnote 2, p.8.

To compute energy expenditure levels, each energy price must be multiplied to its respective level of consumption. Those were retrieved from EIA (2012, p.341) prior to 1950 and then EIA (2016, p.7) from 1949 to 2012. In order to express these energy expenditure levels as fractions (or percentages) of GDP we used the nominal US GDP and deflator estimates of Johnston & Williamson (2016) in continuous year-to-year time series from 1850 to 2012.

**Table 6.1 Sources and original units of recomposed US energy prices, 1850–2012.**

| Energy | Time and spatial coverage | Source | Original unit |
|---|---|---|---|
| Coal | 1850–2012: US average anthracite price. | US Census Bureau (1975a, pp.207–209) from 1850 to 1948; EIA (2012, p.215) from 1949 to 2011; EIA (2013, p.54) for 2012. | Nominal $US/80-lb from 1800 to 1824; then nominal $US/short ton.[1] |
| Oil | 1861–1944: US average;<br>1945–1983: Arabian Light posted at Ras Tanura;<br>1984–2012: Brent dated. | British Petroleum (2015) for the entire period. | Nominal $US/barrel. |
| Gas | 1890–2012: US average price at the wellhead. | US Census Bureau (1975a, pp.582–583) from 1890 to 1915; Manthy (1978, p.111) from 1916 to 1921; EIA (2016, p.145) from 1922 to 2012. | Nominal $US/thousand cubic feet. |
| Electricity | 1907–2012: US average retail price. | US Census Bureau (1975b, p.827) from 1907 to 1959; EIA (2016, p.141) from 1960 to 2012. | Nominal $US cents/kWh. |
| Woodfuel | 1850–2012: US average | Howard & Westby (2013, p.67); all commodities Warren & Pearson (1933, pp.25–27); Manthy (1978, p.90). | Nominal $US/thousand board feet. |
| Biofuels | 2000–2012: US ethanol (E85).<br>2002–2012: US biodiesel (B20). | US Department of Energy (2016) | Nominal $US/Gasoline Gallon Equivalent.[2] |

Data for the World

It is of course quite complicated to estimate the average annual price of a given energy type at the global scale. To be accurate in such estimations, one should formally have all national energy prices and consumption quantities and compute for each year a quantity-weighted average price of each energy. Given the broad time frame of our analysis, such an estimation is simply impossible. Consequently, we will use the different energy prices estimated for the US as global proxies by considering that international markets are competitive and that large spreads between regional energy prices cannot last for long due to arbitrage opportunities. This assumption is fairly relevant for oil and gas. On the other hand, the hypothesis that the average international prices of coal, electricity, woodfuel, and modern biofuels follow their US equivalents is a rather coarse assumption. For instance, in the case of coal, transportation costs over long distances can be very high so that spreads between prices of two different exporting countries have necessarily occurred in the past. Furthermore, by

---

[1] 1 metric tonne = 1000 kg = 1.10231 short ton; 80-lb = 36.29 kg.

[2] 1 Gasoline Gallon Equivalent = 114,100 BTU.

using a single price for coal, we ignore the manifold qualities of coal (from the high energy content of anthracite to the lowest quality of lignite). As our coal price estimate is representative of anthracite, our coal expenditure estimates are probably high estimations of the actual levels of coal expenditure because we surely slightly overestimate the exact quality-weighted global average price of coal. Computing such a quality-weighted global average price of coal would be possible if we knew both the proportions of all the different coal qualities in the total global coal production (i.e. the quality mix of the global coal supply) and their respective prices, for each year between 1850 and 2012. As far as we know, such data is unfortunately not available.

As show in Appendix B global primary energy productions were retrieved from the online data portal of The Shift Project (2015) which is built on the original work of Etemad & Luciani (1991) for 1900–1980 and EIA (2014) for 1981–2012. Prior to 1900, we completed the different fossil fuel time series with the original five-year interval data of Etemad & Luciani (1991) and filled the gaps by linear interpolation. The work of Fernandes et al. (2007) and Smil (2010) was used to retrieve historical global consumption of traditional biomass energy (including woodfuel and crop residues but excluding fodder and traditional windmills and water wheels). The gross world product (GWP) we used comes from Maddison (2007) for 1850 to 1950 and from the GWP per capita of The Maddison Project (2013) multiplied by the United Nations (2015) estimates of the global population for 1950 to 2010. In order to obtain GWP estimates for 2011 and 2012 we used the real GWP growth rate of the World Bank (2016a).

Data for the UK

Regarding the UK, Fouquet (2008; 2011; 2014) has provided a lot of very long-term (1300–2008) data and analyses. More specifically, the prices (£2000/toe[1]) and quantities (Mtoe) of coal, oil, gas, electricity, wood, and fodder consumed in the UK were retrieved from Fouquet (2008) for the period 1300–1699, and we used updated values from Fouquet (2011; 2014) for the period 1700–2008. UK GDP (£2000) was retrieved from Fouquet (2008).

### *6.2.3 Energy Expenditures Estimates for the USA, the UK, and the World*

US energy expenditure

In Figure 6.2a we compare three different estimates of US energy expenditure as a fraction of GDP from 1850 to 2012 (excluding or including wood as marketed energy, and including wood with the total proxy calculation). We also show in this figure the US estimation of King et al. (2015b). Figure 6.2b shows the decomposition of our first best estimate (including wood as marketed energy) by energy type. In Figure 6.3 we relate graphically our first best estimation of the US level of energy expenditure (as a fraction of GDP) to the GDP growth rate from 1951 to 2010.

Quite logically, in early industrial times the US level of energy expenditure was low for fossil energy (coal, oil, and gas) and non-fossil electricity. In 1850 woodfuel expenditure still represented 14% of the US GDP when the overall energy expenditure level was 16%. The low price of coal (cf. Figure 6.1) explains that total energy expenditure decreased from 1850 (16%) to the 1900s (8%) despite a huge increase in consumption. From 1910 to 1945, total energy

---

[1] 1 toe = 1 tonne of oil equivalent = 42 GJ.

expenditure was about 14% of GDP because of ever-increasing (cheap) coal use and the newly increasing consumption of (expensive) hydroelectricity. From 1945 to 1973, which was the period of highest economic growth rates for the US and all other industrialized economies, the level of energy expenditure steadily declined from about 8% to 4%. In 1974 US energy expenditure surged to 10% of GDP, and in 1979 it reached 14.5%. These well-known periods, respectively called the first and second oil crisis, pushed industrialized economies into major recessions. After the beginning of the 1980s, the level of US energy expenditure decreased and reached a minimum of 4.2% in 1998. Then, US energy expenditure rose again (mainly because of the oil price) and reached 7.8% in 2008. After a fall to 5.7% in 2009, US energy expenditure remained around 7% from 2010 to 2012.

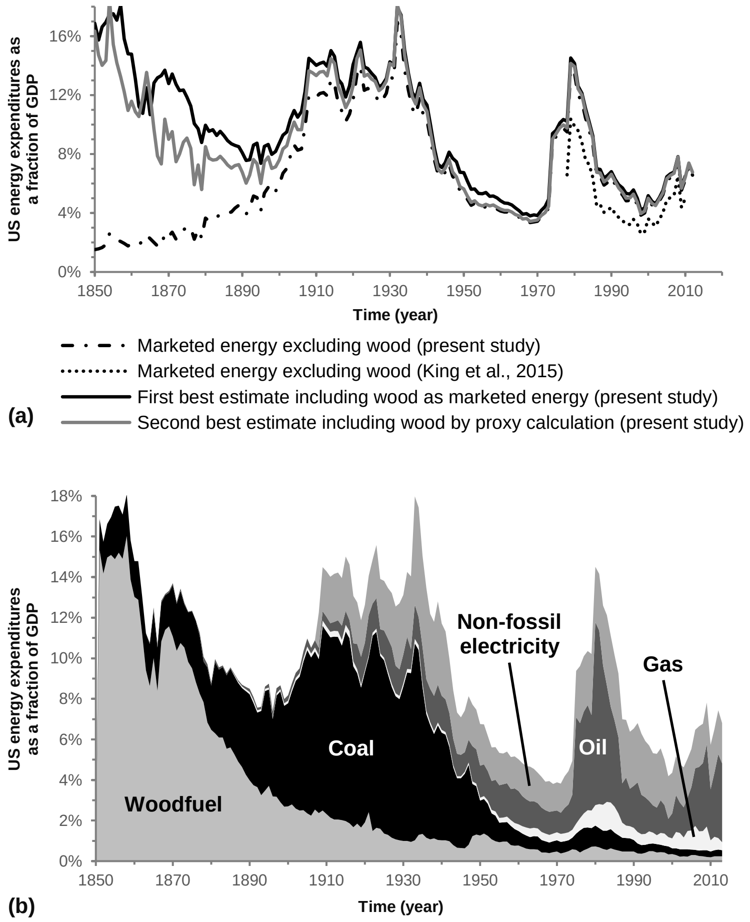


**Figure 6.2 US energy expenditure estimates, 1850–2012.**
**(a) Excluding wood as marketed energy as in King et al. (2015b) vs. including wood as marketed energy vs. total proxy calculation. (b) First best estimate decomposition by energy type.**

Figure 6.2a shows that including traditional biomass energy with the total proxy calculation yields a *second best* estimation of total US energy expenditure that is quite consistent with the *first best* estimation that includes wood as marketed energy. Hence, for a given country for which woodfuel prices are not available, the proxy calculation allows an adequate estimation of the order of magnitude of the total energy expenditure level. Similarly, if consumed quantity estimations of fodder and traditional windmills and water wheels were available without knowing their respective prices, the proxy calculation would be adequate to estimate the actual total level of energy expenditure.

Figure 6.3 indicates that some economic growth recessions are clearly preceded by surges in energy expenditure, and so the importance of energy in such a context cannot be ignored. This is obviously the case for the two oil crisis of the 1970s. On the other hand, the underlying energy basis is harder to discern for some economic recessions. In 1953, for instance, bad monetary policy decisions triggered a demand-driven recession in 1954. In the same way, the 1958 *Eisenhower recession* caused by depressed sales of cars and houses and high interest rates seems disconnected from any energy base.

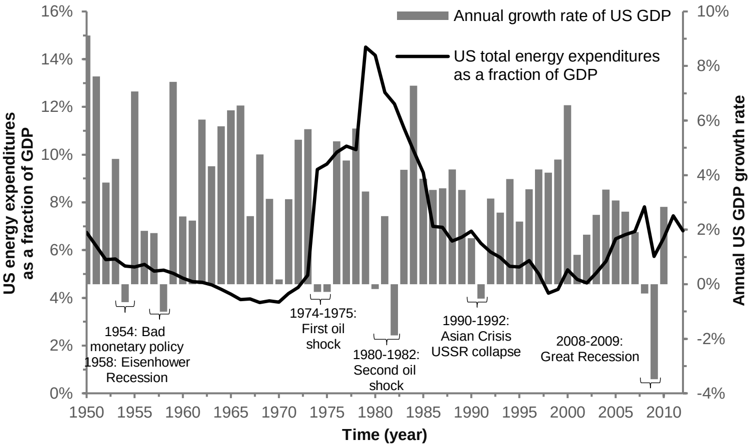


**Figure 6.3 US energy expenditure vs. GDP growth rate, 1951–2010.**

But other political and market-induced economic turmoil are in fact related to energy. For instance, though agonizing and crippled by multiple problems, the oil-exporting USSR probably collapsed in 1990, and not before, because low and ever decreasing oil prices in the early 1990s made its public budget untenable. Similarly, the bursting of the subprime bubble of 2007–2008, which initiated the Great Recession, was in place for a few years and was probably just waiting for a push that rocketing oil prices made visible. Figure 6.3 is only meant to give qualitative intuitions about the energy-economic growth relation but upcoming results will support the main evidence of this chapter: energy is obviously not the only driver of economic growth but it is surely the most recurrent determinant of the economic process.

Global energy expenditure

Figure 6.4a shows our estimation of global energy expenditure as a fraction of GWP from 1850 to 2012 (excluding or including wood as marketed energy, and including wood with the total proxy calculation). This figure also shows the global estimation of King et al. (2015b). Figure 6.4b shows the decomposition by energy type of our global first best estimate including wood as marketed energy.

World results confirm our analysis of the US energy-economy system. Periods of very high energy expenditure relative to GDP (from 1850 to 1945), or surges (in 1973–74 and 1978–79) are associated with low economic growth rates. On the contrary, periods of low or decreasing energy expenditure (from 1945 to 1973) are associated with high and increasing economic growth rates.

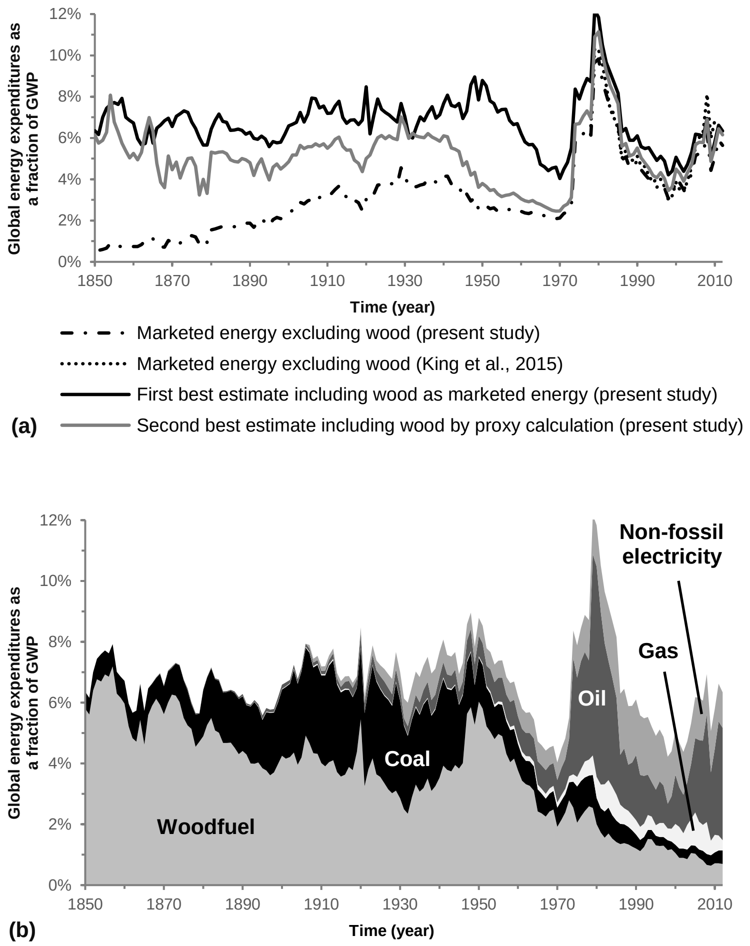


**Figure 6.4** **World energy expenditure estimates, 1850–2012.**
**(a) Excluding wood as marketed energy as in King et al. (2015b) vs. including wood as marketed energy vs. total proxy calculation. (b) First best estimate decomposition by energy type.**

UK energy expenditure

As our results in Figure 6.5 show, when energy expenditure is calculated as far back as 1300, ignoring expenditure related to food (supplied to laborers to obtain power) and fodder (provided to draft animals to obtain power) could lead to a huge underestimation of the past energy cost burden. Indeed, getting total non-human-food energy (but including fodder indispensable to obtain draft animals' power) used to account for 30–40% of the economic product of the UK in the late Middle Ages, and adding human food energy (indispensable to obtain power from laborers) increases such an estimate to 50–70% for the same early times. Even in 1700, food supplied to laborers, wind used for ships and mills, and fodder provided to draft animals accounted for nearly 45% of the total primary energy supply of the UK, and still represented 20% in 1850 (Fouquet 2010). Nevertheless, Figure 6.5 shows that, compared to the US and the global economy (Figures 6.2 and 6.4 respectively), the energy transition of the UK toward fossil fuels was far more advanced in 1850. At that particular time, coal expenditure was about 9.5% of GDP in the UK, but only 2% in the US, and 1.5% at the global scale. Furthermore, ignoring food and fodder as we did for the US and the global economy, the relatively low level of "fossil + woodfuel" energy expenditure of the UK between 1700 and

1800 is, to our mind, a clear sign of the decisive role played by cheap coal to give the UK a head start over other nations in the Industrial Revolution that ultimately lead to the Great Divergence among well-off western and less-developed eastern countries (Pomeranz 2000; Kander et al. 2013; Wrigley 2016).

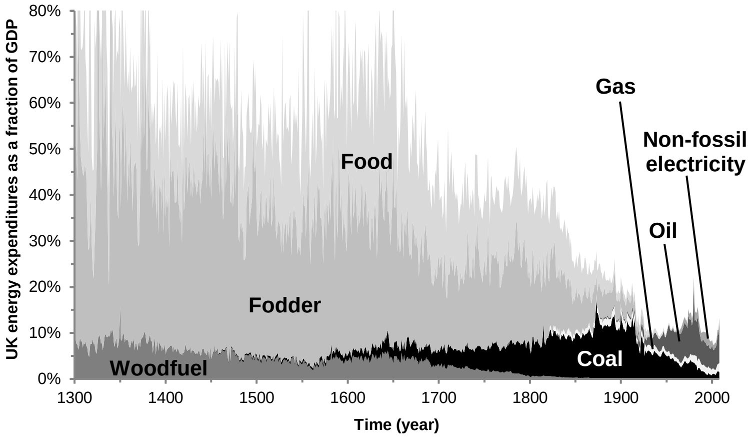


**Figure 6.5 UK energy expenditure estimates, 1300–2008.**

Sensitivity analysis of the US energy expenditure to the GDP data

In Figure 6.6 we test the sensitivity of the US total energy expenditure to the choice of the GDP estimate. As could have been expected, our total energy expenditure estimates are consistent after 1950 since international accounting rules were only established after the Second World War. Before 1950, nominal GDP estimates and deflator estimates vary more widely among authors but without generating excessive differences in our energy expenditure estimates.

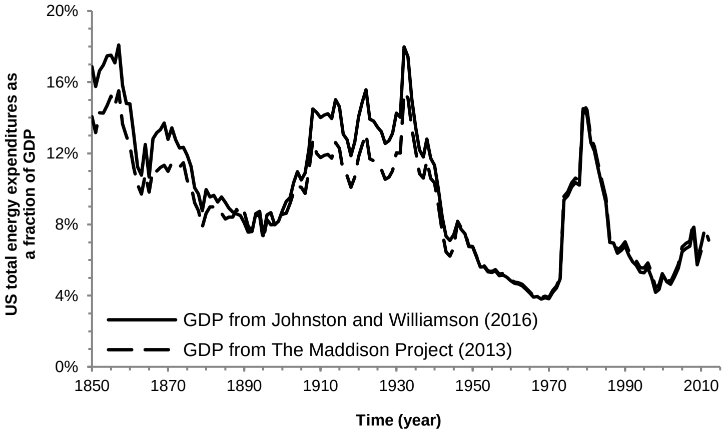


**Figure 6.6 Sensitivity analysis of US energy expenditure to GDP estimate, 1850–2012.**

Consistency with Bashmakov's "first law"

According to our results, it seems that the "first energy transition law" postulated by Bashmakov (2007) concerning the stability of energy costs to income ratios ("with just a limited sustainable fluctuation range") is valid for the post-Second World War era but not for earlier periods. On the whole, our results suggest that the ratio of US energy expenditure to GDP has decreased from an average value of 11% for the period 1850–1950 to a lower average value of 5.7% for 1950–2012. The fact that Bashmakov's "first law" does not hold in the very long-term is even more visible if we observe the energy requirements of the UK between 1300 and 2008 in Figure 6.5.

## 6.3 ENERGY EXPENDITURE AS A LIMIT TO GROWTH

In this section we investigate quantitatively the hypothesis that energy expenditure can be a limit to growth. We first present the methodology used to estimate the maximum level of energy expenditure above which economic growth cannot be positive (from a statistical point of view). Then, we show how to translate this result into the maximum tolerable price of energy, or in other words, the minimum required EROI of society. Although we are pretty confident in using US energy prices as global proxies for estimating the global level of energy expenditure, the following equations and econometric tests will only be applied to the US due to the lack of availability, consistency, and confidence that we have in global estimates of population and capital formation (as a fraction of GWP). Indeed, continuous population estimates are readily available for the US for the entire period of study of this article, whereas continuous estimates of global population are only available since 1950. Regarding gross capital formation as a fraction of GDP, the World Bank (2016b) proposes estimates from 1960 to 2013 for the US, but only from 1970 to 2013 for the global economy. Moreover, confidence in data is logically higher for a well-administered nation like the US than for global estimates.

### *6.3.1 US Economic Growth Regressions*

In the US case, once total expenditure of marketed energy ($X_{total\ marketed}$) is computed, we can perform different multivariate linear regressions. The US GDP growth rate (obtained from Johnston & Williamson 2016) representing the dependent variable can be regressed on several explanatory variables, namely: energy expenditure as a fraction of GDP (in which all marketed energy forms can be considered, or just a subset such as oil), capital formation as a fraction of GDP (retrieved from the World Bank 2016b), and the US population (from Johnston & Williamson 2016). As we suspect population to be a poor proxy for labor availability, we will also test in our regressions the explanatory power of the US unemployment rate (provided by the US Bureau of Labor Statistics 2016). The general formula of the multivariate linear regression we study is:

$$\frac{\dot{GDP}}{GDP} = \alpha + \theta_1 \frac{X_{total\ marketed}}{GDP} + \theta_2 \frac{Capital\ formation}{GDP} + \theta_3 \Delta Population + \varepsilon. \qquad (6.6)$$

Where $\frac{\dot{GDP}}{GDP}$ is the US economic growth rate, $\alpha$ is the intercept, $\theta_1$ (for which we logically anticipate a negative value because, the higher the energy expenditure, the lower the discretional consumption and investment to fuel growth) represents the sensitivity of the economic growth rate to the level of energy expenditure as a fraction of GDP, $\theta_2$ is the sensitivity of the economic growth rate to the capital formation as a fraction of GDP, and $\theta_3$ is the sensitivity of the economic growth rate to population first difference $\Delta Population$. It is important to point out that the main advantage of our approach is that it takes into account both the impact of energy prices and energy efficiency on economic growth. Indeed, it should be remembered that energy expenditure as a fraction of GDP can be broken down as the average price of energy times the energy intensity (inverse of energy efficiency) of the economy:

$$\frac{X_{total\ marketed}}{GDP} = \frac{\sum_j P_j E_j}{GDP} = P_{average} \times \frac{\sum_j E_j}{GDP} = P_{average} EI. \qquad (6.7)$$

Where $EI$ is the energy intensity of the economy. So, rather than considering only the impact of energy price or energy quantity fluctuations on economic growth, as is usually done in econometric studies, we suppose here that energy prices impact the economy variously depending on the energy efficiency of the economy. The higher the energy intensity of the economy, the higher the negative impact of energy price increases.

Table 6.2 gives the results of the different ordinary least square (OLS) regressions we have performed following equation (6.6) where US economic growth is the dependent variable and US energy expenditure, US capital formation, US population first difference, and US unemployment rate are the different explanatory variables.[1] In specification (I) we have considered only oil expenditure, capital investment, and US population. As suspected, population seems to be a poor proxy for labor as its effect is not statistically significant. To correct for this shortcoming, we introduce the US unemployment rate in all other specifications (II to V). Therefore specification (II) is similar to specification (I) except for the labor proxy. Specification (III) takes into account all three fossil energies (coal, oil, and gas), capital investment, and unemployment rate. In specification (IV) energy expenditure includes all fossil energies, non-fossil electricity and wood, whereas specification (V) is the same as (IV) with additional dummies to control for the impact of peculiar events, namely the two oil shocks (1974 and 1979), the oil counter-shock (1986), and the global Great Recession (2009).

We found a statistically significant (most of the time at 1% level) decreasing relation between the US economic growth and the level of energy expenditure as a fraction of GDP between 1960 and 2010 for all specifications. Increasing energy expenditure as a fraction of GDP is a sufficient condition for a decline in US economic growth but this factor is not a necessary condition for a contraction of the economy since geopolitical, institutional, socioeconomic, and climatic events, and the unavailability of capital and labor can also reduce economic growth. Specification (II) shows that an increase of one percentage point in oil expenditure is correlated to a 0.60 decrease in US economic growth. When all fossil fuel expenditure (III), or all energy expenditure (IV) are taken into account instead of just oil, energy expenditure still has a statistically significant negative impact on economic growth, but the correlation is slightly weaker. An increase of one percentage point in fossil (respectively total) energy expenditure is statistically correlated to a 0.55 (respectively 0.48) decline in US economic growth. As shown by specification (V), this result is robust to the inclusion of several

---

[1] Results of unit root tests performed for all time series can be found in Appendix K.

dummy variables in order to control for the impact of particular events. Capital investment is always positively significant at 1% level. Each point of investment as a fraction of GDP raises economic growth by slightly more than one percentage point.

**Table 6.2 Results of multivariate regressions for the US economy, 1960–2010.**

| | Dependent variable: US GDP growth rate | | | | |
|---|---|---|---|---|---|
| Specification | **(I)** | **(II)** | **(III)** | **(IV)** | **(V)** |
| Constant | -0.180740<br>(0.045554)*** | -0.260034<br>(0.052281)*** | -0.277873<br>(0.052875)*** | -0.276934<br>(0.053082)*** | -0.264372<br>(0.057749)*** |
| US oil expenditure | -0.406652<br>(-3.294917)*** | -0.608737<br>(0.131068)*** | | | |
| US fossil energy expenditure | | | -0.554234<br>(0.118643)*** | | |
| US total energy expenditure including wood | | | | -0.475930<br>(0.114248)*** | -0.522700<br>(0.152441)*** |
| US capital investment | 0.957723<br>(0.206976)*** | 1.206830<br>(0.205298)*** | 1.288255<br>(0.208538)*** | 1.307545<br>(0.208708)*** | 1.238985<br>(0.223166)*** |
| US population first difference | -1.15E-09<br>(8.48E-10) | | | | |
| US unemployment rate | | 0.434847<br>(0.252110)* | 0.522721<br>(0.257490)** | 0.605045<br>(0.284391)** | 0.724169<br>(0.334816)** |
| dum1974 | | | | | -0.018473<br>(0.004933)*** |
| dum1979 | | | | | 0.011897<br>(0.010671) |
| dum1986 | | | | | -0.017128<br>(0.006443)** |
| dum2009 | | | | | -0.031794<br>(0.011243)*** |
| $R^2$ | 0.493143 | 0.533416 | 0.540681 | 0.520744 | 0.583032 |
| $R^2$ Adjusted | 0.460790 | 0.503634 | 0.511362 | 0.490154 | 0.515154 |
| **Residual tests** | | | | | |
| Durbin-Watson | 1.683556 | 1.744475 | 1.765262 | 1.687059 | 1.623818 |
| White | 2.150983** | 3.467135*** | 3.462751*** | 3.514716*** | 1.900220* |
| Arch (1) | 0.170333 | 2.75E-05 | 0.025367 | 0.034475 | 0.006592 |
| Jarque-Bera | 0.686598 | 0.454564 | 0.305434 | 0.409832 | 4.152342 |
| Shapiro-Wilk | 0.986928 | 0.971674 | 0.972468 | 0.983263 | 0.968714 |
| CUSUM test | Stability: yes | Stability: yes | Stability: yes | Stability: yes | / |
| CUSUM squared test | Stability: yes | Stability: yes | Stability: yes | Stability: yes | / |

Note: Robust standard error estimates are reported in parentheses. * Significant at 10% level, ** 5% level, ***1% level.

Surprisingly, the US unemployment rate is positively correlated with economic growth when the impact of energy expenditure and capital investment is also taken into account. To check this result, we made a simple regression of US economic growth on the US unemployment rate and found the classic decreasing relationship. Moreover, when we perform univariate linear regressions of the unemployment rate on capital formation (as a fraction of GDP) and on energy expenditure (as a fraction of GDP), we find that the unemployment rate is positively correlated to energy expenditure (the higher the energy expenditure as a fraction of GDP, the higher the unemployment rate) and negatively correlated to capital investment (the higher the capital investment as a fraction of GDP, the lower the unemployment rate). These results indicate that the apparently strange positive correlation between economic growth and unemployment is not caused by a flaw in our data or methodology. The residual checks converge toward the assumption of normality of residuals and the absence of autocorrelation, although there is some evidence for the presence of heteroscedasticity, thus we use robust standard error. The CUSUM and CUSUM squared tests indicate that the estimated coefficients are stable overtime.

It is worth noting that performing the same multivariate linear regressions at the global scale yields very similar results, in particular the statistically significant negative correlation between energy expenditure and economic growth. We choose not to reproduce these results because the CUSUM and CUSUM squared tests indicate that the estimated coefficients are not stable overtime for this global approach.

Regarding the diverse econometric regressions performed in this chapter, an alternative approach might be to analyze the relationship between energy expenditure (as a fraction of GDP) and the growth rate of per capita GDP instead of total GDP as we did. We tested this option and found similar outcomes. We deliberately choose to focus our study on GDP growth and not per capita GDP growth in order to remain consistent with the existing literature.

We could also suppose the existence of threshold effects in the relationship between economic growth and energy expenditure (as a fraction of GDP) instead of the linear relationship assumed in this chapter. This assumption is a key point of Bashmakov's work (2007). Whether this relationship is linear or not (threshold existence) involves the presence or absence of trade-offs between high energy expenditure as a fraction of GDP (causing high effort of energy efficiency) and high economic growth. Unfortunately, considering the restricted number of observations (fewer than ten) that we have for high levels of energy expenditure, it remains quite complicated to derive robust econometric estimations for such high regimes. The use of panel data could be a good way to overcome this technical barrier, and this option might be explored in further work.

### *6.3.2 Maximum Tolerable Level of Energy Expenditure/Energy Price and Minimum EROI*

Maximum tolerable level of energy expenditure

Using equation (6.6), it is easy to find the particular value of US energy expenditure (as a fraction of GDP) that leads to zero economic growth. In other words, we can define the maximum level of energy expenditure (as a fraction of GDP) above which positive economic growth is impossible. We call $\beta_{total}$ this maximum level of energy expenditure, with:

$$\beta_{total} = \left(\frac{X_{total\,marketed}}{GDP}\right)_{max} = \frac{-\alpha - \theta_2 \dfrac{Capital\,formation}{GDP} - \theta_3 \Delta Population}{\theta_1}. \quad (6.8)$$

Following equation (6.8), and replacing parameters $\alpha, \theta_1, \theta_{2,} \theta_3$ by the estimated values of specification (IV) (so respectively, -0.28, -0.48, 1.31, and 0.61), and the mean values of capital formation as a fraction of GDP (0.2244) and unemployment rate (0.0598), we find the central value of the maximum tolerable level of total energy expenditure:

$$\beta_{total} = \frac{0.28 - 1.31 \times 0.2244 - 0.61 \times 0.0598}{-0.48} = 0.11. \tag{6.9}$$

Using a Wald test, we can provide a minimum and maximum $\beta_{total}$ at 5% level. We find that $0.09 < \beta_{total} < 0.131$. This result means that, in the US, if the fraction of energy expenditure is higher than 11% of GDP (with a 95% confidence interval of [9%–13.1%]), economic growth is statistically lower than or equal to zero (all others variables being equal to their mean values). Using parameter values from specification (II), we can perform the same test for oil expenditure only and derive the maximum tolerable level of oil expenditure for the US economy, $\beta_{oil}$, which is equal to 6% (with a 95% confident interval of [4.6%–7.5%]). Our results support the qualitative suppositions advanced by Murphy & Hall (2011ab) and Lambert et al. (2014).

Maximum tolerable quantity-weighted average price of energy

Defining the maximum level of energy expenditure above which positive economic growth is impossible can be reformulated as the maximum aggregated price of marketed energy $P_{average\,max}$ that the economy can tolerate to still present a slightly positive growth rate. Of course, this hypothetical maximum tolerable price of aggregated energy depends on the energy intensity of the US economy as shown here in (6.10):

$$P_{average\,max} = \frac{\beta_{total}}{\frac{E_{total\,marketed}}{GDP}}. \tag{6.10}$$

Using the US estimations of $\beta_{total}$ and $\beta_{oil}$ calculated in the previous section, we can easily compute the maximum price of aggregated energy, $P_{average\,max}$, and the maximum price of oil, $P_{oil\,max}$, above which US economic growth should statistically become negative. Obviously, both estimates are absolutely not static but time dependent since for any given year, they respectively depend on the current total energy intensity and the current oil intensity of the US economy:

$$P_{average\,max,t} = \frac{\beta_{total}}{\frac{E_{total\,marketed,t}}{GDP_t}} = \frac{0.11}{\frac{E_{total\,marketed,t}}{GDP_t}}, \tag{6.11}$$

$$P_{oil\ max,t} = \frac{\beta_{oil}}{\frac{E_{oil,t}}{GDP_t}} = \frac{0.06}{\frac{E_{oil,t}}{GDP_t}}. \tag{6.12}$$

Relation (6.12) describing the maximum tolerable price of oil for the US economy as a function of its oil intensity is represented in Figure 6.7 and compared with the actual historical course of the oil price between 1960 and 2012. We could have easily drawn this figure for total aggregated energy, but given the importance of oil for the US economy, we think that focusing on the oil price is more advisable here. If we consider the last data point of the econometric estimation we have for year 2010, Figure 6.7 indicates that the price of oil would have had to reach 16977 $1990/TJ (equivalent to 173 $2010 per barrel) instead of its real historical value of 8315 $1990/TJ (84 $2010 per barrel), to annihilate US economic growth. Figure 6.7 also shows that in 2008 the oil price was pretty close to the limits to growth zone, and one must not forget that average annual values are not representative of extremes and potentially lasting events. Oil prices increased continuously in the first half of 2008 reaching 149 $2010 on July 11. This supports the idea that the surge in oil expenditure at this time indeed played a limits to growth role in lowering discretionary consumption and hence revealing the insolvency of numerous US households. A preliminary additional mechanism is to consider that instabilities on the financial market in 2007 led numerous non-commercial agents to take positions on apparently more reliable primary commodities markets (Hache & Lantz 2013). This move inevitably puts upward pressure on prices, and in particular the oil price, which increased energy expenditure as a fraction of GDP to the point of triggering a limit-to-growth effect. Similarly, from 1979 to 1982, the actual oil price was above or slightly below its maximum tolerable value, which explains that US economic growth had very little chance of being positive during those years. On the contrary, at the time of the oil counter-shock of the late 1980s, the oil price was four times below its maximum tolerable level, so that the oil expenditure constraint was very loose at this time.

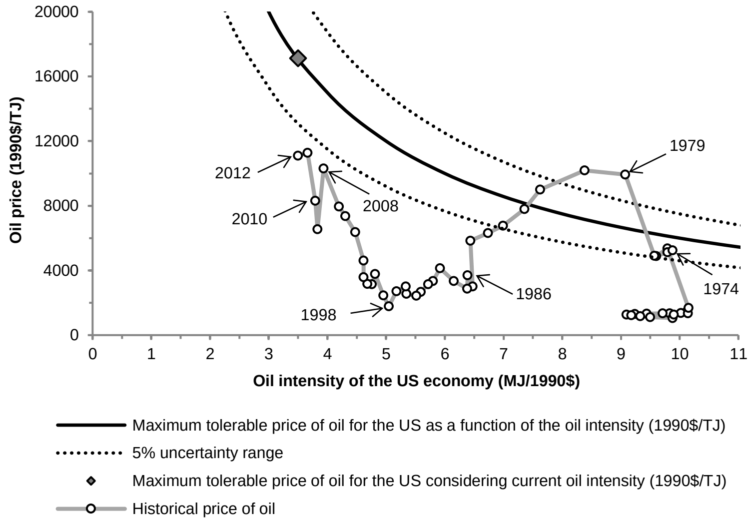


**Figure 6.7 Maximum tolerable US price of oil as a function of the economy oil intensity.**

### Minimum EROI required to enjoy positive economic growth

As presented in Section 4.1.1, King et al. (2015a) point out that the classical EROI definition is rather loose and that a clear distinction should be made between yearly *power return ratios* (PRRs) of annual energy flows and *energy return ratios* (ERRs) of full life cycle energy systems (i.e. cumulated energy production divided by total lifetime invested energy) which more formally represent EROIs. Understandably, energy return ratios represent integrals of power return ratios over the entire life cycle of the energy system under consideration. Following King & Hall (2011), an estimate of the annual or yearly EROI of a given economy (taking into account only marketed energies for which prices are available) can be expressed as a function of the quantity-weighted average price of aggregated marketed energy, $P_{average}$, the average monetary-return-on-investment (MROI) of the energy sector (i.e. its gross margin), the gross domestic product (GDP), and the total supply of marketed energy $E_{total\ marketed}$:

$$EROI = \frac{MROI}{P_{average} \times \frac{E_{total\ marketed}}{GWP}}. \tag{6.13}$$

If we replace $P_{average}$ in (6.13) by the expression (6.10) of $P_{average\ max}$, we obtain an expression of the $EROI_{min}$, which is the minimum societal EROI that the energy system must have in order for the economy to enjoy a positive rate of growth:

$$EROI_{min} = \frac{MROI}{\beta_{total}}. \tag{6.14}$$

As can be seen in equation (6.14), two variables are needed to calculate the minimum aggregated EROI, $EROI_{min}$, required for having positive economic growth in the US: the maximum tolerable level of energy expenditure $\beta_{total}$ previously calculated, and the average monetary-return-on-investment (MROI) of the energy sector. In Section 4.2.2 such average MROI of the US energy sector is estimated at the value of 1.158±0.02 for 1850–2012 (meaning that between 1850 and 2012, on average, the gross margin of the US energy sector has been about 15.8%, with a standard deviation of 2%). Using this average value of 1.158 for the MROI, and the value of 0.11 previously calculated for $\beta_{total}$, we estimate that the US economy requires a primary energy system with an $EROI_{min}$ of 11:1 in order to enjoy a positive rate of growth. Taking the uncertainty range (at 5% level) of $\beta_{total}$ ([0.09–0.131]), and considering an MROI varying between 1.05 and 1.2, the sensitivity of the $EROI_{min}$ ranges from 8:1 to 13.5:1.

To the best of our knowledge, there are only three studies that discuss potential values for minimum societal EROI. Hall et al. (2009) offer a technical minimum EROI of 3:1 for oil at the well-head. These authors postulate (without explicit calculation) that a higher value of 5:1 would be necessary to just support our current complex societies, but that a minimum EROI around 12–14:1 is probably necessary to sustain modern forms of culture and leisure. Weißbach et al. (2013) give a minimum required EROI of 7:1 for OECD countries without a clear explanation of the underlying calculation. Finally, the study by Lambert et al. (2014), based on simple (although nonlinear) correlations between EROI and the Human Development Index (HDI) in cross sectional data, arrive at a minimum required societal EROI in the range 15–25:1 for contemporary human societies.[1]

Now that we have estimated that, at current energy intensity, the US requires a minimum societal EROI of 11:1 (with a most likely interval[2] of [8–13.5]) in order to possibly have

---

[1] In their study Lambert et al. (2014) define a minimum EROI in order to reach a minimum HDI which is quite different from our minimum EROI below which positive economic growth is statistically compromised.

[2] This expression is used because it is impossible to formally define a 5% or 10% confidence interval for the $EROI_{min}$. Indeed, such confidence interval is known for $\beta_{total}$, but not for the MROI for which only a standard deviation of 2% is known. Hence,

positive economic growth, the temptation is to compare this value to the representative EROI of different energy systems in order to assess their *growth-compatibility*. Such a comparison appears rather perilous. First, studies proposing EROI sometimes calculate ratios of annual gross energy produced to annual energy invested which hence represent power return ratios (PRRs) or annual energy return ratios (ERRs) comparable to our $EROI_{min}$; but more formally, EROIs should describe ratios of cumulated energy production to total energy invested, and such estimates can be found in the literature too. Second, there is no such thing as an *average representative EROI value* for a given energy system. Each energy system has a particular EROI that depends on the considered input boundary (Murphy et al. 2011). The bottom line is that the order of magnitude of net energy ratios (be it ERRs or PRRs) are important, precise calculated values are not. Hence, the different numbers given here must absolutely be understood as representative orders of magnitude. Coal, oil, and gas have respective representative EROIs of about 80–100:1, 20–30:1, and 40–60:1. Hydropower projects have high EROIs of about 50–100:1 (but the global remaining hydro potential will probably come to saturation in a few decades). New renewable technologies toward which human future is destined have relatively lower EROIs, with average values for wind power, photovoltaic panels, and first generation biofuels respectively around 15–20:1, 4–6:1, and 1–2:1 (Hall et al. 2014). Adding the intermittent nature of renewable energy to this perspective suggests that (so far) new renewable technologies hardly seem capable of coping with the minimum required societal EROI of 11:1 that we have calculated.

### *6.3.3 Granger Causality between Energy and GDP in the USA from 1960 to 2010*

The last part or our work consists in studying the temporal causality between US energy expenditure (as a fraction of GDP) and US GDP growth rates between 1960 and 2010. There are many causality tests based on different definitions of causality, but the main idea of the Granger (1969) causality test is to verify that adding past data of variable $X_1$ to past data of variable $Y$ enhances the prediction of present values of variable $Y$. If the residuals generated from a model with variable $Y$ and its past only, and from another model with the past of variable $Y$ and the past of variable $X_1$ are significantly different, we can reject the assumption of non-causality from $X_1$ to $Y$ and accept the assumption of a causality running from $X_1$ to $Y$. Formally, it consists in running the following Wald test:

$$H_0\colon \forall\, i \in [1,\dots,k], \theta_{1,i} = 0 \; and \; H_1\colon \exists\, i \in [1,\dots,k], \theta_{1,i} \neq 0,$$

$$Y_t = c + \sum_{i=1}^{i=k} \delta_i Y_{t-i} + \sum_{i=1}^{i=k} \theta_{1,i} X_{1,t-i} + \sum_{i=1}^{i=k} \theta_{2,i} X_{2,t-i} + \sum_{i=1}^{i=k} \theta_{3,i} X_{3,t-i} + \varepsilon_t. \qquad (6.15)$$

---

the interval [8–13.5] was computed to simply get an idea of the sensitivity of the estimated average $EROI_{min}$ but this interval must surely not be taken as a formal result.

We also test the assumption that all the $X_j$ variables are not Granger causing the variable $Y$ by testing $H_0: \forall\, i \in [1, \ldots, k], \theta_{1,i} = \theta_{2,i} = \theta_{3,i} = 0$, and $H_1: \exists\, i \in [1, \ldots, k] \cup j \in [1, \ldots, 3]$, $\theta_{j,i} \neq 0$.

Over the period 1960–2010 for which we have uninterrupted year-to-year data, we performed Granger causality tests to identify the direction of the possible causal relation between the US level of oil expenditure as a fraction of GDP, US capital formation as a fraction of GDP, US unemployment rate, and the growth rate of the US GDP. Our results, presented in Table 6.3, show that we can reject at 5% level the assumption that the level of oil expenditure as a fraction of GDP does not Granger cause economic growth. For the reverse relation, the assumption that growth does not Granger cause the level of oil expenditure (as a fraction of GDP) cannot be rejected at 5% level. In summary, these tests indicate a one way causality from energy expenditure to economic growth at 5% level. Applying the same methodology, we also find a one way causality running from the US level of oil expenditure to the US unemployment rate (Figure 6.8). Finally, the Granger causality test also tends to confirm a feedback relationship between the US economic growth and the US unemployment rate at 5% level. Furthermore, contrary to our static econometric results ( of Table 6.2), the impulse response functions estimated from the vector autoregression (VAR) used in Granger causality tests show in a dynamic way how a variable can be impacted by a modification of another variable. We found that an increase in energy expenditure (as a fraction of GDP) in a given year leads to an increase in the unemployment rate two years later and a decrease in economic growth in the three years following the initial rise in energy expenditure. Quite logically, we observed also that economic growth reacts negatively to a rise in the unemployment rate and positively to a rise in capital investment (as a fraction of GDP).

**Table 6.3 Results of Granger causality tests for the US economy, 1960–2010.**

| **Dependent variable** | **Sources of causation (independent variables) with 1 lag** | | | | |
|---|---|---|---|---|---|
| | Oil expenditure | GDP growth | Unemployment rate | Capital formation | All |
| Oil expenditure | - | 2.321782 | 0.278008 | 0.514794 | 3.049061 |
| GDP growth | 11.61990*** | - | 19.58885*** | 1.083957 | 25.73877*** |
| Unemployment rate | 10.22715*** | 10.69602*** | - | 0.100274 | 46.42257*** |
| Capital formation | 1.243340 | 6.466733** | 9.453183*** | - | 21.49198*** |

Note: To determine the lag order, we used the lag order chosen by the majority of information criteria (in our case 4 out of 5 information criteria indicated an optimal order of one lag). We also checked that the VAR is well specified and that there was no persistent autocorrelation. *corresponds to the F-statistic result of the Fisher test rejecting the assumption $H_0$: "the variable $X_i$ does not Granger cause the variable $Y$" with a 10% risk level, ** 5% risk level, *** 1% risk level.

It is worth adding that using total energy expenditure instead of oil expenditure in the same Granger causality tests yields identical results. However, with those data, autocorrelation problems could only be solved by increasing the number of lags in our relations. Considering the low number of observations that we have, this strategy reduces the robustness of these results and we consequently chose not to reproduce them here.

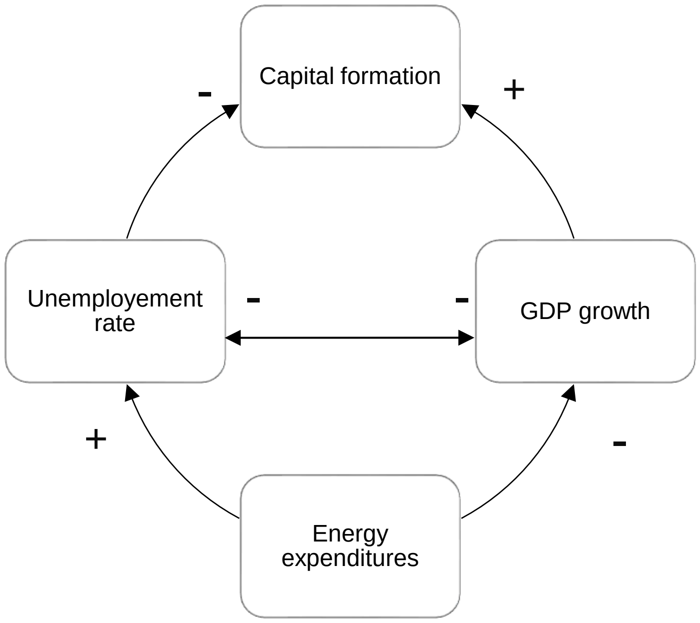


**Figure 6.8 Relationships highlighted by our VAR for the US, 1960–2010.**

In this sixth chapter we estimated the level of energy expenditure, i.e. the amount of GDP diverted to obtain energy, from 1850 to 2012 for the US and the global economy, and from 1300 to 2008 for the UK. Results indicate that the level of energy expenditure in the economy seems to play a limit-to-growth role since as long as it has remained above 6–8% of GDP, high economic growth rates have never occurred in the last several centuries for the US, the UK, and even the global economy. More precisely, periods of high or suddenly increasing energy expenditure levels are associated with low economic growth rates: for instance from 1850 to 1945 (very high energy expenditure levels), from 1975 to 1976 (surge), and from 1981 to 1983 (surge). On the contrary, periods of low and decreasing energy expenditure are associated with high and increasing economic growth rates: for instance from 1945 to 1973, and in the early 2000s.

Furthermore, we were able to show that in order to have a positive growth rate, from a statistical point of view, the US economy cannot afford to allocate more than 11% of its GDP to primary energy expenditure (in the absence of other major limits of a geographical, geopolitical, or institutional nature). This means that considering its current energy intensity, the US economy needs to have at least a societal $EROI_{min}$ of approximately 11:1 (that conversely corresponds to a maximum tolerable average price of energy of twice the current level) in order to present positive rates of growth.

Finally, over the more restricted period 1960–2010 for which we have continuous year-to-year data for the US, we performed several Granger causality tests that consistently show a one way negative causality running from the level of energy expenditure (as a fraction of GDP) to economic growth.

# CHAPTER 7

# LONG-TERM ENDOGENOUS ECONOMIC GROWTH AND ENERGY TRANSITIONS

*“Anyone who believes exponential growth can go on forever*
*in a finite world is either a madman or an economist.”*

Kenneth Boulding

Several studies have focused on the transition between a nonrenewable and a renewable natural resource in a neoclassical analytical framework, but none of them refers to biophysical concepts such as exergy and EROI. Some of these studies (Jouvet & Schumacher 2012; Hartley et al. 2016) are not able to represent simultaneous use of nonrenewable and renewable energy but only successive regimes that use specifically one of the energy forms, which logically generates some energy crisis behavior at the time of the abrupt switch. Conversely, the optimal growth model of Tahvonen & Salo (2001) is able to represent for an abstract economy a first phase of economic development that relies exclusively on renewable energy, a second phase where renewable and nonrenewable energy are simultaneously used, and a third phase where the share of nonrenewable energy decreases because of increasing extraction costs, thus leading to a society that relies on renewable energy only. In Tsur & Zemel (2005) the attention is more focused on the R&D investments that allow a reduction in the cost of use of a backstop technology, but the broader effect of an energy transition on economic growth is not studied. Acemoglu et al. (2012) have studied the importance of the substitutability between nonrenewable and renewable inputs in directing endogenous technical change, and the influence of the optimal mix of environmental policies between carbon tax and R&D subsidy. In light of what has been presented in chapters 3 to 6, it is clear that there is a need to build a bridge between the different literatures related to the endogenous economic growth theory, the biophysical economics perspective, and the transition between nonrenewable and renewable energy forms. It is the very purpose of this last chapter to bridge this theoretical gap.

## 7.1 STRUCTURE OF THE MODEL

### *7.1.1 Goal, Inspiration, and Economic Product Allocation of the Model*

Goal

The goal of the present chapter is to propose a theoretical model of long-term endogenous economic growth that takes into account the underlying physical reality of the economic system. The theoretical positioning of our work and its background literature have already been presented in previous chapters. In the remainder of the present section, we present the model of a decentralized economy in which the accessibility of primary nonrenewable and renewable exergy, and the efficiency with which those inputs are converted into useful exergy services, determine the production of a final output good that is consumed or saved to allow investment. In Section 7.2 we specify the calibration procedure to global historical data and show that the model adequately reproduce, from 1750 to 2010, the pattern of historical global energy production, technological change, and economic growth. We then run simulations of the model in order to study its dynamics in future times, in particular we assess the necessary conditions for a smooth transition towards an almost-renewable-only regime. We analyze in Section 7.3 the interest of the implementation of a price on the polluting emissions of nonrenewable energy in order to smooth the transition towards increasing renewable energy in an original simulation setting in which the energy transition has negative impacts on economic growth. We then conclude our work and discuss some of our hypotheses for further research developments.

Inspiration

The model presented in this chapter builds on Fagnart & Germain (2014) who included the EROI concept in a theoretical model to investigate the possibility of a smooth transition from nonrenewable to renewable energy and its impact on the EROI and economic growth. In this model, uncalibrated simulations can only be done with an initial economy just before the nonrenewable energy peak and no production of renewable energy (which is thus not representative of reality). Hence, despite its novelty, there are different features of this model that we would like to address in the present chapter, namely that (i) the nonrenewable energy is extracted without any capital requirement and consequently presents an infinite EROI, (ii) the backstop technology has a constant capital requirement per unit of energy output, (iii) technological change is bounded but completely exogenous, and (iv) the production function in the final good sector is of the Leontief type. In order to address these particular settings and others mentioned earlier, we provide an endogenous economic growth model subject to the physical limits of the real world, meaning that nonrenewable and renewable energy production costs have functional forms that respect physical constraints, and that technological change is precisely defined as gains in the aggregate efficiency of primary-to-useful exergy conversion.

Economic product allocation of the model

At each period $t$, the representative household receives the entire macroeconomic income made of the rents from total capital $K_t$ loaned at price $v_t$ and the different profits $\Pi_t$,

$\Omega_t, \Psi_t$ of the respective nonrenewable energy (NRE), renewable energy (RE), and final good sectors. This total income is logically equal to the macroeconomic product $Y_t$, so

$$Y_t = v_t K_t + \Pi_t + \Omega_t + \Psi_t, \qquad \forall\, t \in \{0, \dots, T\}. \tag{7.1}$$

The capital stock of the economy $K_t$ should not be considered as pure physical capital but rather as *labor activated effective capital services* since we do not represent the population and labor dynamics. *Labor activated* means that the capital services should be understood as the output result of the aggregation (in a production function that we do not detail) of pure physical capital with routine labor hours provided by the population. *Effective* means that the capital services output also contains some human capital in the form of skills and hand-eye coordination (to which the recent contribution of information and communication technologies should be added).

Given that we wish to calibrate the model on global historical data for the period 1750–2010, and then pursue simulation up to the point where nonrenewable energy is almost not used, we assume for simplicity a unitary depreciation rate of capital, implying that the time period $t_{length}$ between $t$ and $t+1$ corresponds to the average capital lifetime set to 20 years. As a consequence, it is acceptable to not represent any maximization behavior of the intertemporal welfare of the households, and rather to consider that the representative household consumes from the macroeconomic output $Y_t$ what is left over after the investment $I_t$ has been fulfilled. Hence, with $C_t$ representing the discretionary consumption at the macroeconomic level, we have

$$Y_t = C_t + I_t, \quad \forall\, t \in \{0, \dots, T\}. \tag{7.2}$$

This means that the capital services cost is in fact constant and worth$v \equiv (1+\mu)^{t_{length}}/\lambda$, where $\mu$ corresponds to the annual real interest rate of the economy, and $\lambda > 0$ represents the productivity of the transformation of investment goods into productive capital.[1] The dynamics of the capital investment level is

$$I_t = \frac{K_{t+1}}{\lambda}. \tag{7.3}$$

Furthermore, equilibrium on the capital market must hold at each time period. Hence, the total capital stock of the economy $K_t$ is the sum of the NRE sector capital $Z_t$, the RE sector capital $G_t$, and the final good sector $H_t$.

$$K_t = \begin{cases} Z_t + G_t + H_t, & \forall\, t \in \{0, \dots, T_e\} \\ G_t + H_t, & \forall\, t \in \{T_e + 1, \dots, T\}. \end{cases} \tag{7.4}$$

---

[1] Introducing the intertemporal welfare optimization behavior of the representative household implies a non-constant capital cost. This fact greatly complicates the calibration procedure but given the time frame chosen for the simulation it only smooth the results without changing any of the qualitative outcomes of the model.

Where $T_e$ is the final time period of nonrenewable energy resource exploitation.

### *7.1.2 Profit Maximization of NRE, RE, and Final Good Producers*

Profit maximization of the NRE producer

The ultimately recoverable resource (URR), $\mathcal{R}$, represents the total amount of accessible primary nonrenewable energy in the Earth underground and exploited during the length of time $T_e$ by a representative price-taking firm. It is assumed that the representative firm does not know $\mathcal{R}$ but observes that its production cost evolves as the nonrenewable resource is progressively depleted. Extracting the gross primary nonrenewable energy quantity $R_t$ implies consuming some capital services $Z_t$. Furthermore, a fraction $0 < \chi_{NRE} < 1$ of the gross primary production $R_t$ is self-consumed by the NRE sector. Accordingly, in each period $t$ the NRE producer chooses an amount of capital services $Z_t$ in order to supply the quantity $R_t(1-\chi_{NRE})$ of available primary nonrenewable energy to the final good sector at the unitary price $p_t$. Hence, the producer solves

$$\max_{R_t,Z_t} \Pi_t = (1-\chi_{NRE})p_t R_t - vZ_t\,, \qquad \forall\, t \in \{0,\dots,T_e\} \tag{7.5}$$

under constraint,

$$Z_t = (R_t D_t)^{\frac{1}{\theta}}\,, \qquad with\; 0 < \theta < 1 \;\; \forall\, t \in \{0,\dots,T_e\}. \tag{7.6}$$

Where $D_t$ represents the capital intensiveness of the extraction process (i.e. the capital requirement per unit of gross primary NRE output), whose detailed definition is given in Section 7.1.3. The fact that $0 < \theta < 1$ means that returns to scale are decreasing in the NRE sector. Recalling that $t_{length}$ is the time period length in real years between *t* and *t+1*, we have

$$\sum_{t=0}^{T_e} t_{length} R_t \leq \mathcal{R}\,, \qquad \forall\, t \in \{0,\dots,T_e\}, \tag{7.7}$$

and,

$$\lim_{T_e \to +\infty} R_t = 0. \tag{7.8}$$

After the insertion of (7.6) into (7.5), the first order condition with respect to $R_t$ gives

$$R_t = \left[\frac{p_t(1-\chi_{NRE})\theta}{D_t^{\frac{1}{\theta}} v}\right]^{\frac{\theta}{1-\theta}}, \qquad \forall\, t \in \{0,\dots,T_e\}. \tag{7.9}$$

Profit maximization of the RE producer

We suppose that a very large (and never binding) flow of renewable primary energy (aggregation of solar radiant energy, geothermal and tidal energies) is accessible to the economy and that a price-taking representative firm is in charge of its exploitation. In order to capture the gross primary renewable energy flow $F_t$ some capital $G_t$ is obviously necessary and a fraction $\chi_{RE}$ of the gross energy output is self-consumed. Thus, in each period $t$, the RE producer maximizes its profit $\Omega_t$ and consequently chooses a capital stock $G_t$ in order to deliver the flow $F_t(1-\chi_{RE})$ of available primary renewable energy sold at the unitary price $p_t$ by solving

$$\max_{F_t,G_t}\Omega_t = (1-\chi_{RE})p_tF_t - vG_t\ , \qquad \forall\, t\in\{0,\dots,T\} \tag{7.10}$$

under constraint,

$$G_t = (F_tB_t)^{\frac{1}{\gamma}}, \qquad with\; 0<\gamma<1 \;\; \forall\, t\in\{0,\dots,T\} \tag{7.11}$$

Where $B_t$ represents the capital intensiveness of the RE producer (i.e. the capital requirement per unit of RE output), whose detailed definition is given in Section 7.1.3. The fact that $0 < \gamma < 1$ means that returns to scale are decreasing and that consequently the capital intensiveness of the RE firm increases with the production level.[1] Once (7.11) is injected into (7.10), the first order condition with respect to $F_t$ leads to,

$$F_t = \left[\frac{p_t(1-\chi_{RE})\gamma}{B_t^{\frac{1}{\gamma}}v}\right]^{\frac{\gamma}{1-\gamma}}, \quad \forall\, t\in\{0,\dots,T\}. \tag{7.12}$$

Profit maximization of the final good producer

The total primary energy $E_t$ available to the final good sector is

$$E_t = \begin{cases} R_t(1-\chi_{NRE}) + F_t(1-\chi_{RE}), & \forall\, t\in\{0,\dots,T_e\} \\ F_t(1-\chi_{RE}), & \forall\, t\in\{T_e+1,\dots,T\}. \end{cases} \tag{7.13}$$

This available primary energy $E_t$ is combined with capital services $H_t$ in a Cobb-Douglas production function in order to produce the final output good $Y_t$ representing the macroeconomic product. The formulation of a production function must be independent of the choice of units, hence we introduce the dimensionless variables $y_t \equiv \frac{Y_t}{Y_0}$, $a_t \equiv \frac{A_t}{A_0}$, $e_t \equiv \frac{E_t}{E_0}$, and, $h_t \equiv \frac{H_t}{H_0}$, where $Y_0$, $A_0$, $E_0$, and $H_0$ are given quantities in the initial reference period. Hence,

---

[1] As shown at the end of Appendix G, Dale et al. (2011) used two databases of the National Renewable Energy Laboratory (NREL 2010a; 2010b) to demonstrate that the frequency of wind and solar power sites in the USA is an inverse function of their productive potential, meaning that over time the availability of optimal sites will decrease. In the same way, Hoogwijk et al. (2004) and Honnery & Moriarty (2009) have shown that as wind energy production increases, the marginal capacity factor of wind turbines decreases.

$$y[a,e,h]_t = (a_t e_t)^{\alpha} {h_t}^{1-\alpha}, \quad \forall\, t \in \{0, \dots, T\}. \tag{7.14}$$

and,

$$Y_t = y[a,e,h]_t\, Y_0, \quad \forall\, t \in \{0, \dots, T\}. \tag{7.15}$$

The output elasticities of useful energy and capital services inputs are constant and respectively represented by $\alpha$ and $1-\alpha$. We follow Ayres & War (2009) and other authors such as Cleveland et al. (1984) who have earlier emphasized that the aggregate technology level of the economy is formally represented by the efficiency with which primary exergy contained in fossil fuels and solar flow is converted into useful exergy services in the forms of light, heat, electricity, and mechanical drive. Hence, technological change corresponds to *gains in the aggregate efficiency of primary-to-useful exergy conversion* (i.e. increases of $A_t$). Formally, the primary-to-useful exergy conversion efficiency $A_t$ is the product of: (i) the primary-to-final efficiency with which primary exergy contained in fossil fuels and the solar flow is converted into final exergy in the forms of carriers such as liquid fuels (e.g. gasoline), compressed gas, and high-temperature heat; with (ii) the final-to-useful efficiency with which exergy contained in these final forms is converted into useful exergy services in the forms of light, heat, and mechanical drive. Hence, $A_t E_t$ represents useful exergy services[1] provided in the forms of light, heat, electricity and mechanical drive to the real economy. Of course, another part (smaller to our mind) of the improvement of economic productivity comes from the division and organization of labor, the enhancements of laborer skills, the beneficial effects of inclusive institutions (which, for example, protect private property rights and consequently incentivize innovation and R&D), and the recent contribution of information and communication technologies. Such attributes of human capital are embedded in the labor activated effective capital services $H_t$, whose optimal value is found by considering the final good price as the numeraire, and that the representative firm in the final good sector seeks to solve

$$\max_{E_t, H_t} \Psi_t = Y_t - p_t E_t - v H_t, \quad \forall\, t \in \{0, \dots, T\} \tag{7.16}$$

under constraint (7.14) and (7.15). The resolution of this problem implies combining the first order conditions with respect to $E_t$ and $H_t$ in order to find

$$H_t = \left[\frac{1-\alpha}{\alpha}\frac{p_t}{v}\right] E_t, \quad \forall\, t \in \{0, \dots, T\}. \tag{7.17}$$

Combining (7.17) with (7.14)-(7.15) in the first order condition with respect to $E_t$ gives (after mathematical arrangements)

---

[1] The term *useful work* is also used in the literature (see Ayres & Warr 2009; Warr et al. 2010; Warr & Ayres 2012).

$$p_t = \left[\alpha \frac{Y_0}{H_0}\left(\frac{A_t H_0}{A_0 E_0}\right)^{\alpha}\left(\frac{1-\alpha}{\alpha v}\right)^{1-\alpha}\right]^{\frac{1}{\alpha}}, \quad \forall\, t \in \{0, \dots, T\}. \tag{7.18}$$

For the clarity of the remainder of the presentation let us define now the saving rate of the economy $S_t$ as the ratio of investment $I_t$ to the macroeconomic product $Y_t$:

$$S_t = \frac{I_t}{Y_t}, \quad \forall\, t \in \{0, \dots, T\}. \tag{7.19}$$

### *7.1.3 Endogenous Technological Change, Unitary Capital Requirements, and EROI*

Endogenous technological change

The technological level $A_t$ is necessarily bounded from above by a strictly positive constant $\overline{A}$ representing the maximum efficiency of primary-to-useful exergy conversion that the economy will ultimately reach in the future. This positive upper bound is strictly inferior to one since the second law of thermodynamics imposes that perfect (i.e. 100%) efficiency of primary-to-useful exergy conversion is impossible.[1] This uncertain parameter (for which we test several values later) is taken as exogenous. The technological level increases over time at speed $\xi_t$ and at some point (when the maximum limit $\overline{A}$ is close) the incremental gains in $A_t$ are so small that the dynamic system describing the economy is in a quasi-steady state. Hence, with $t_{\Delta A_t\, max}$ as the particular time at which the growth rate of the technological level (i.e. the technological change) is maximum, we define the following law of motion for $A_t$ as

$$A_t = \underline{A} + \frac{\overline{A} - \underline{A}}{1 + \exp\left(-\xi_t\left(t - t_{\Delta A_t\, max}\right)\right)}, \qquad \forall\, t \in \{0, \dots, T\}. \tag{7.20}$$

Furthermore, we suppose that the speed of convergence $\xi_t$ between the initial technological level $\underline{A}$ and its asymptotic value $\overline{A}$ (verifying $0 < \underline{A} < \overline{A}$) depends on the variation of the knowledge stock of the economy. This (potentially infinite) knowledge stock (that we do no represent) depends on the effort deployed in the R&D sector in previous periods that itself follows the saving rate of the economy (i.e. the level of investment compared to the level of economic production) of these same previous periods. In addition, the more recent the saving rate, the higher its influence on $\xi_t$. Hence, we define the speed of convergence $\xi_t$ of the technological level as the first order exponential smoothing of the saving rate of the economy

---

[1] As already expounded in Section 3.3.3, Ayres & Warr (2009, p.52–53) highlight that technological change at the macro level is ultimately defined by the limiting efficiency of all metallurgical, chemical and electronic processes at micro levels, which in turn depends essentially on the properties of structural materials. Indeed, some technologies, such as prime movers and many metallurgical reduction and synthesis processes, depend on the temperatures, and in some cases, pressures, achievable in a confined space. These are limited by the strength and corrosion resistance (chemical inertness) of structural materials at elevated temperatures. In the same way, turbine efficiencies also depend on the precision with which blades, piston rings, gears and bearings can be manufactured, which depends in turn on the properties of materials being shaped and the properties of the ultra-hard materials used in the cutting and shaping of tools.

during the $N$ previous periods (where $N$ is defined through calibration). With $\sigma$ as the share of the macroeconomic investment going to R&D, we have

$$\xi_t = \begin{cases} \sigma S_0 & for\ t = 0 \\ \sigma\left[\left(\frac{2}{N+1}\right)S_{t-1} + \left(1 - \frac{2}{N+1}\right)\xi_{t-1}\right], & \forall\, t \in \{1, \dots, T\}. \end{cases} \tag{7.21}$$

This formulation of the technological level insures that in our model both technological change and economic growth are endogenous.

Unitary capital requirement in NRE sector

An accurate formulation of the nonrenewable capital cost $D_t$ should, to our mind, necessarily reproduce three facts: (i) the cost associated with nonrenewable energy extraction must necessarily increase with cumulative production, this is because easier-to-exploit resources are used up first before attention turns to deeper and more remote resources[1] (see Murphy & Hall (2011b) for a graphic representation of this fact in the case of oil production); (ii) the initial unitary cost of NRE production was above the RE production cost before the nineteenth century, this is necessary to explain that despite being known since antiquity, coal was not produced on an industrial scale before wood charcoal became scarce and expensive in England in the late eighteenth century; (iii) learning processes and R&D have so far allowed a decrease of the NRE production cost. Since we did not find in the literature a formulation that would suit these three prerequisites, the NRE unitary capital requirement proposed in this model is (to the best of our knowledge) unique to the present article.

The capital requirement per output unit of nonrenewable energy, $D_t$, is composed of two parts as defined in equation (7.22) and shown in Figure 7.1. The first part increases through the extraction process because of the quality depletion of the NRE resource, and the second part decreases through time thanks to learning and R&D processes. Hence, the first term depends on the ratio of nonrenewable resource depletion $\phi_t$, varying between 0 when the nonrenewable energy resource is still virgin and 1 when it is fully depleted. The second term depends on the ratio of technological level advancement, varying between 0 when the technological level equals its lower bound $\underline{A}$ and 1 when the technological level equals its upper bound $\overline{A}$. The idea behind this relation is that even though we do not explicitly represent the specific R&D of the energy sector, we can fairly assume that the different sectors of the economy evolve with *technological consistency*.[2] Hence, even if from a formal point of view $A_t$ represents the efficiency of primary-to-useful exergy conversion in the final good sector, we postulate that this variable, after being normalized between 0 and 1, is a proxy of the technological level of the energy sector.

---

[1] The rational tendency of humans to first use easier-to-exploit high quality resources before turning towards harder-to-exploit lower quality resources is commonly known as the *Best First Principle* (Hall & Klitgaard 2012).

[2] Of course, some sectors might have faster technological improvements than others, and that is particularly true regarding the distinction we make between the two primary energy-producing sectors on the one hand, and the final good energy-consuming sector on the other. Nevertheless, we think that on average such sectoral discrepancies in technological levels cannot last more than a few decades. Indeed, on a larger time horizon, technological level gaps between sectors would imply investment opportunities and subsequent reallocation of financial capital and hence R&D. Considering that our time step is twenty years, we think that postulating a technological consistency between the different sectors of our model is rather justified.

With $D_0$ as the initial capital cost per NRE output unit, $\widetilde{D}$ as the maximum capital cost reduction thanks to learning and R&D processes, and $\delta$ as a constant parameter representing the rate of quality degradation of the NRE resource, we can define $D_t(\phi_t, A_t)$ as

$$D_t(\phi_t, A_t) = D_0 exp^{\delta \phi_t{}^{\omega_1}} - \widetilde{D}\left(\frac{A_t - \underline{A}}{\overline{A} - \underline{A}}\right)^{\omega_2}, \qquad \forall\, t \in \{0, \dots, T_e\}. \tag{7.22}$$

Where $\omega_1$and $\omega_2$ are positive constants determined when calibrating the model on historical global data in section 4. The exploited resource ratio $\phi_{NRE,t}$ is defined as

$$\phi_t = \frac{t_{length} \sum_{i=0}^{t-1} R_i}{\mathcal{R}} \in [0,1], \qquad \forall\, t \in \{0, \dots, T_e\}. \tag{7.23}$$

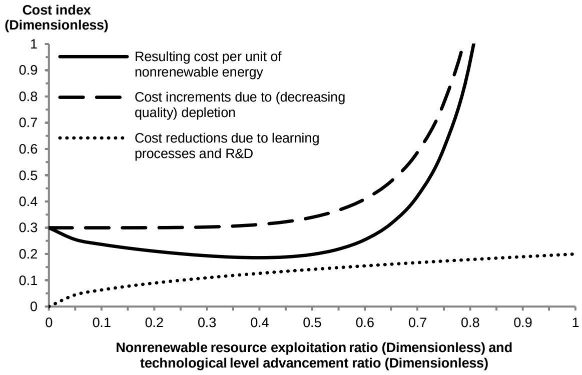


**Figure 7.1 Capital cost per output unit of nonrenewable energy.**
**This example is obtained with $D_0 = 0.3$, $\widetilde{D} = 0.2$, $\omega_1 = 5$, $\omega_2 = 0.5$, and $\delta = 4$.**

Unitary capital requirement in RE sector

To be accurate, the capital requirement per renewable energy output unit, $B_t$, should be represented by a decreasing function since, over time, less capital is necessary to capture the same amount of primary renewable energy thanks to learning processes and R&D. Furthermore, as for the NRE sector, we postulate that the RE sector is *technologically consistent* with the rest of the economy, so that $B_t$ is a function of $A_t$. The sigmoid decreasing function $B_t(A_t)$ describing the capital cost per unit of renewable energy output starts at value $\overline{B}$ and decreases at a constant speed $\tau > 0$ to a strictly positive bound $\underline{B}$ since the production of any RE flow would always require a minimum quantity of capital,

$$B_t(A_t) = \overline{B} - \frac{\overline{B} - \underline{B}}{1 + \exp\left(-\tau\left(A_t - A_{\Delta B_t max}\right)\right)}, \qquad \forall\, t \in \{0, \dots, T\}. \tag{7.24}$$

Where $A_{\Delta B_t max}$ is the particular technological level at which the function $B_t$ presents an inflexion point (i.e. the rate of degrowth of $B_t$ is maximum when $A_t = A_{\Delta B_t max}$). In addition, we suppose that the final unitary cost of renewable energy production $\underline{B}$ also depends on the final technological level $\overline{A}$ of the final good sector. Precisely, the higher the ratio of technological level gain $\overline{A}/\underline{A}$, the lower the final unitary cost of RE production $\underline{B}$ should be compared to its initial value $\overline{B}$. Hence, with $\eta$ as a parameter found through calibration to historical data, we suppose

$$\underline{B} = \frac{\overline{B}}{\left(\overline{A}/\underline{A}\right)^{\eta}}, \qquad with\ 0 < \eta < 1. \tag{7.25}$$

EROI of energy sectors

In order to define the energy-return-on-investment (EROI) of the two energy sectors, we need to breakdown the saving rate $S_t$ into three parts $S_{H,t}$, $S_{Z,t}$, $S_{G,t}$defined respectively as the fraction of the economic output of period *t* invested in the final good sector (respectively NRE, RE sector) in period $t+1$.

$$S_t = S_{H,t} + S_{Z,t} + S_{G,t},\ \ with\ \ S_{H,t} = \frac{H_{t+1}}{\lambda Y_t}, \qquad S_{Z,t} = \frac{Z_{t+1}}{\lambda Y_t}, and\ \ S_{G,t} = \frac{G_{t+1}}{\lambda Y_t}. \tag{7.26}$$

According to Hall et al. (2014), the EROI is “the ratio between the energy delivered by a particular fuel to society and the energy invested in the capture and delivery of this energy”. King et al. (2015a) point out that this definition is rather loose and that a clear distinction should be made between yearly *power return ratios* (PRRs) of annual energy flows and *energy return ratios* (ERRs) of full life cycle energy systems (i.e. cumulated energy production divided by total energy invested) which more formally represent EROIs. Understandably, energy return ratios represent integrals of power return ratios over the entire life cycle of the energy system under consideration. Recall that in our theoretical model defined in discrete times, the duration between two consecutive periods equals the capital services lifetime. As a consequence, PRRs and ERRs are exactly the same in our particular theoretical setting. Furthermore, PRRs and ERRs can differ regarding the system boundary of their energy output (numerator) and energy input (denominator). We will consider as energy outputs the production levels of gross primary energy $R_t$ or $F_t$ in the NRE or RE cases respectively. The invested energy usually takes two forms: *direct* energy inputs in the form of self-consumption and external energy investments (generally as final carriers like electricity or liquid fuels), and *indirect* inputs energy embodied in capital and services. Since the energy sectors of the model represent upstream sectors producing primary energy, and considering that we do not represent downstream sectors that convert primary energy into final forms, the direct energy inputs of the two primary energy producing sectors are only represented by their respective self-consumptions. In order to calculate the indirect energy investments embodied in capital services, let us consider the example of the nonrenewable sector in which the production of the gross primary energy output $R_t$ requires the capital stock level $Z_t$ that comes from the fraction $S_{Z,t-1}$ of economic output $Y_{t-1}$. Since the production of $Y_{t-1}$ has required the consumption of the primary energy $E_{t-1}$, it follows that the quantity of indirect energy embodied in the NRE sector capital services used in period *t* is $S_{Z,t-1}E_{t-1}$. Finally, given all previous precisions and referring to King et al.

(2015a) definitions, the EROI (a denomination we keep for convenience) we compute is formally a gross power ratio that due to our discrete time setting equals its integral over time (i.e. gross energy ratio ). For the NRE sector, the $EROI_{NRE,t}$ is defined as

$$EROI_{NRE,t} = \frac{R_t}{R_t \chi_{NRE} + S_{Z,t-1} E_{t-1}}. \tag{7.27}$$

Similarly, the $EROI_{RE,t}$, of the RE production in period *t* is

$$EROI_{RE,t} = \frac{F_t}{F_t \chi_{RE} + S_{G,t-1} E_{t-1}}. \tag{7.28}$$

Finally, it is possible to define the EROI of the whole primary energy sector since delivering the total gross primary energy $R_t + F_T$ to the final sector has directly required the self-consumption $R_t \chi_{NRE}$ and $F_t \chi_{RE}$, and indirectly required the embodied energy $S_{Z,t-1} E_{t-1} + S_{G,t-1} E_{t-1}$. Thus globally, the $EROI_t$ of the entire energy sector in period *t* is

$$EROI_t = \frac{R_t + F_t}{R_t \chi_{NRE} + F_t \chi_{RE} + (S_{Z,t-1} + S_{G,t-1}) E_{t-1}}. \tag{7.29}$$

## 7.2 CALIBRATION AND SIMULATIONS

### *7.2.1 Global Historical Data*

Four time series are used to calibrate the model on global historical data: nonrenewable primary exergy production, renewable primary exergy production, efficiency of primary-to-useful exergy conversion, and GWP. Since the time period $t_{length}$ between $t$ and $t+1$ corresponds to the average capital lifetime set to 20 years, our historical data time series consist of fourteen discrete points from 1750 ($t = 0$) to 2010 ($t = 13$). Of course, the different references used to retrieve global historical data do not always propose values for the specific year we need. Hence, the 20–years interval estimations of global historical data reported in Table 7.1 are rounded up values and comparison with the respective data references are provided in Figure 7.2.

To suit the model structure we have aggregated in a single NRE production the different historical data for global primary production of coal, oil, gas, and nuclear energy. Following Kümmel (2011) we made the assumption that these primary energy forms expressed in exajoules per year (EJ/yr, where 1 EJ = $10^{18}$ J) represent 100% exergy. Energy production values have been retrieved through the online data portal of The Shift Project (2015), which is built on the original work of Etemad & Luciani (1991) for the 1900–1980 time period and EIA (2014) for 1981–2010. Prior to 1900, we have completed the different fossil fuel time series with the original 5-year interval data of Etemad & Luciani (1991) and filled the gaps using linear interpolation. In the same way, the historical global primary production of biomass energy (woodfuel and crop residues[1]) from Fernandes et al. (2007) and Smil (2010) were aggregated with the historical global renewable energy production of The Shift Project (2015) for hydro, wind, solar, geothermal, wastes, ocean (wave, tidal, OTEC), and modern biofuels into a single primary renewable energy production expressed in EJ/yr.

**Table 7.1 20–years interval historical estimates used for model calibration, 1750–2010.**

| Time period (actual year) | Nonrenewable primary exergy production (EJ/year) | Renewable primary exergy production (EJ/year) | Efficiency of primary-to-useful exergy conversion (dimensionless) | Gross world product (Billion Int. G-K. $1990/year) |
|---|---|---|---|---|
| 0 (1750) | 0.00 | 19.55 | 0.0250 | 435 |
| 1 (1770) | 0.05 | 19.65 | 0.0250 | 465 |
| 2 (1790) | 0.20 | 19.85 | 0.0255 | 495 |
| 3 (1810) | 0.55 | 20.50 | 0.0265 | 530 |
| 4 (1830) | 1.00 | 21.25 | 0.0278 | 765 |
| 5 (1850) | 2.20 | 22.05 | 0.0300 | 920 |
| 6 (1870) | 6.00 | 22.75 | 0.0320 | 1115 |
| 7 (1890) | 14.70 | 22.95 | 0.0360 | 1675 |
| 8 (1910) | 31.50 | 23.20 | 0.0420 | 2550 |
| 9 (1930) | 42.50 | 25.50 | 0.0510 | 3720 |
| 10 (1950) | 70.30 | 28.00 | 0.0650 | 5315 |
| 11 (1970) | 201.5 | 38.35 | 0.0800 | 13720 |
| 12 (1990) | 326.5 | 52.30 | 0.1000 | 27350 |
| 13 (2010) | 480.0 | 74.25 | 0.1250 | 54150 |

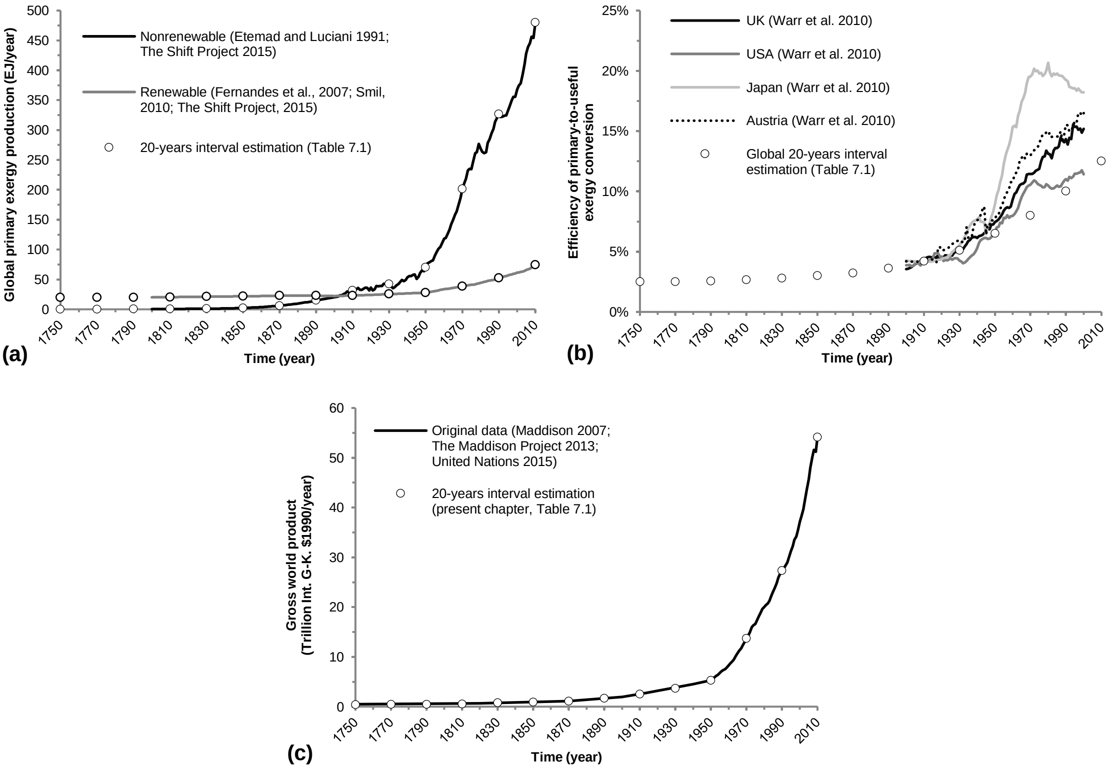


**Figure 7.2 Original data and 20-years interval estimates for model calibration, 1750–2010.**
**(a) nonrenewable and renewable global primary exergy production, (b) efficiency of primary-to-useful exergy conversion, (c) gross world product.**

---

[1] Formally, fodder supplied to draft animals should be added to traditional biomass energy estimates, but it is generally discarded due to difficulties of estimation. This is also the case for traditional windmills and water wheels.

We give in Table 7.1 an estimation of the efficiency of primary-to-useful exergy conversion from 1750 to 2010 at global scale. It is important to emphasize that this estimate does not come from any calculation but only represents a *best guess* after considering the work of Warr et al. (2010), who have estimated the efficiency of primary-to-useful exergy conversion for the US, the UK, Japan, and Austria from 1900 to 2000 as shown in Figure 7.2b.

Regarding the gross world product (GWP) expressed in Billion 1990 International Geary-Khamis dollars,[1] we use the data of Maddison (2007) from 1750 to 1949 and the GWP per capita of The Maddison Project (2013) multiplied by the United Nations (2015) estimates of global population from 1950 to 2011.

### *7.2.2 Parameters, Scenarios, and Calibration*

Parameters and scenarios

All simulations are performed up to the time horizon $T = 25$, which corresponds to the year 2250. The initial technological level is logically set to $A_0 = \underline{A} = 0.025$ and we can also easily define $Y_0 = 435$ (Billion G-K. \$1990) from Table 7.1. Parameters $\chi_{NRE}$ and $\chi_{RE}$ are arbitrarily set equal to 0.01 because we have no reliable data to choose otherwise. All other parameter values summarized for clarity in Table 7.2 are necessarily found through the calibration of the model to historical data. We have performed such procedures with two prerogatives: (i) the calibration must remain robust under the different scenarios that are tested; (ii) the scenarios must differ by the least possible number of differences in parameter values. Logically we found that the main determinant of a given scenario is the ultimate value $\overline{A}$ towards which technological level $A_t$ converges. As shown in Figure A1b of the Appendix it can be fairly assessed that the global technological level $A_t$ has evolved from 0.025 in 1750 to 0.125 in 2010. Since the maximum *attainability* of the efficiency of primary-to-useful exergy conversion is necessarily below 1, we have tested several values between 0.15 and 0.95 and decided to present the results for four scenarios, respectively described by the following $\overline{A}$ values: 0.25, 0.35, 0.45 and 0.65. Once $\overline{A}$ is defined, we found that in order to respect the objectives (i) and (ii) previously cited, only two additional parameters needed to be tuned, namely $\widetilde{D}$ and $t_{\Delta A_t\,max}$. Hence, the four different scenarios, called *Low*, *Medium*, *High*, and *Extra-High*, are exactly determined by their common parameters synthetized in Table 7.2, and their specific parameters presented in Table 7.3.

One important parameter of the model merits specific attention: the nonrenewable energy ultimately recoverable resource (URR),[2] noted $\mathcal{R}$. This parameter represents the total

---

[1] It will be recalled that the 1990 International Geary–Khamis dollar (Int. G-K. \$1990), more commonly known as the international dollar, is a standardized and fictive unit of currency that has the same purchasing power parity as the U.S. dollar had in the United States in 1990.

[2] According to British Petroleum (2015): the “URR is an estimate of the total amount of a given resource that will ever be recovered and produced. It is a subjective estimate in the face of only partial information. Whilst some consider URR to be fixed by geology and the laws of physics, in practice estimates of URR continue to be increased as knowledge grows, technology advances and economics change. The ultimately recoverable resource is typically broken down into three main categories: cumulative production, discovered reserves and undiscovered resource”. On the other hand, Sorrell et al. (2010) highlight that unlike reserves, URR estimates are not dependent on technology assumptions and thus should only be determined by geologic hypotheses. Unfortunately, this apparent contradiction of the URR definition is only a tiny example of the fuzziness of points of view that one could find in the literature regarding the different notions of nonrenewable resources and reserves.

amount of nonrenewable energy that may be recovered at positive net energy yield, i.e. at EROI greater or equal to unity. To obtain the value of the aggregated fossil URR we use the recent work of McGlade & Ekins (2015) ant take their best estimates for oil (Gb: giga barrel), gas (Tcm: terra cubic meters), and coal (Gt: giga tonnes), which for the record are in accordance with the last IIASA Global Energy Assessment report (IIASA 2012). For uranium (EJ: exajoule = $10^{18}$ J), we aggregate the best estimate of conventional and unconventional uranium resource provided by the IIASA (2012), giving the rounded value of 14,500 EJ. After conversion and aggregation, the total nonrenewable URR value retained for our simulations is 177,500 EJ as can be seen in Table 7.4. Contrary to what one might think, sensitivity analyses of the model to this parameter (not presented in the chapter due to lack of space) have shown a great robustness of its qualitative results. Changing the value of $\mathcal{R}$ does not change the dynamics of the model because it is necessarily balanced by a change in other parameter values (in particular $D_0$, $\omega_1$, and $\omega_2$) in order for the calibration to the historical data to remain valid.

**Table 7.2 Specific parameters values of *Low*, *Medium*, *High*, and *Extra-High* scenarios.**

| Parameter | Definition (unit) | *Low* | *Medium* | *High* | *Extra-High* |
|---|---|---|---|---|---|
| $\overline{A}$ | Final technological level of the economy, i.e. final efficiency of primary-to-useful exergy conversion in the final good sector (dmnl). | 0.25 | 0.35 | 0.45 | 0.65 |
| $t_{\Delta A_t\, max}$ | Time of maximum technological change (model time period/*actual year*). | 13.35<br>*(2017)* | 14.45<br>*(2039)* | 15.15<br>*(2053)* | 16.0<br>*(2070)* |
| $\widetilde{D}$ | Maximum capital cost reduction per unit of nonrenewable energy thanks to learning processes and R&D (B$/EJ). | 6.180 | 6.295 | 6.365 | 6.458 |

**Table 7.3 Set of parameter values that are common to all scenarios.**

| Parameter | Definition (unit) | Value | Units |
|---|---|---|---|
| $T$ | Time horizon of the model. | 25 | dmnl |
| $t_{length}$ | Time period length in real years between $t$ and $t+1$. | 20 | years |
| $\lambda$ | Transformation productivity of investment goods. | 7.25 | dmnl |
| $\mu$ | Annual real interest rate of the economy. | 0.03 | dmnl |
| $v$ | Constant capital cost (dmnl), with $v \equiv (1+\mu)^{t_{length}}/\lambda$ | 0.249 | dmnl |
| $\alpha$ | Constant output elasticity of useful exergy. | 0.6 | dmnl |
| $\sigma$ | Share of the macroeconomic investment going to R&D. | 0.9 | dmnl |
| $N$ | Number of time periods used to smooth the saving rate of the economy in $\xi_t$. | 4.0 | dmnl |
| $\underline{A}, A_0$ | Initial technological level. | 0.025 | dmnl |
| $\mathcal{R}$ | Ultimately recoverable resource of nonrenewable energy. | 177,500 | EJ |
| $D_0$ | Initial unitary capital cost of NRE production. | 6.35 | B$/EJ |
| $\delta$ | Rate of quality degradation of the NRE resource. | 0.225 | dmnl |
| $\omega_1$ | Power exponent of the ratio of exploited resource $\phi_{NRE,t}$ in the cost increasing part. | 1.05 | dmnl |
| $\omega_2$ | Power exponent of the ratio of exploited resource $\phi_{NRE,t}$ in the cost decreasing part. | 0.05 | dmnl |
| $\overline{B}$ | Initial production cost per unit of renewable energy output. | 1.35 | B$/EJ |
| $\tau$ | Growth rate of $B_t$ towards $B_2$. | 15 | dmnl |
| $\eta$ | Constant used to link the final capital cost of RE production $B_2$ to its initial value $B_1$, and to the technological level gain ratio $\overline{A}/\underline{A}$. | 0.25 | dmnl |
| $\theta$ | Returns to scale in the NRE sector. | 0.5 | dmnl |
| $\gamma$ | Returns to scale in the RE sector. | 0.5 | dmnl |
| $\chi_{NRE}$ | Share of gross primary energy production self-consumed by the NRE sector. | 0.01 | dmnl |
| $\chi_{RE}$ | Share of gross primary energy production self-consumed by the RE sector. | 0.01 | dmnl |
| $\mathrm{H}_0$ | Initial (1750) capital in the final sector. | 745 | B$ |
| $\mathrm{Y}_0$ | Initial (1750) gross world product. | 435 | B$/yr |
| $\mathrm{S}_0$ | Initial (1750) saving rate of the economy. | 0.5 | dmnl |

**Table 7.4 Global ultimately recoverable resource estimates for coal, oil, gas and uranium.
Sources: IIASA (2012), McGlade & Ekins (2015).**

| Energy resource | Global URR (diverse units) | Conversion factors (diverse units) | Global URR$^*$ (EJ) |
|---|---|---|---|
| Coal | 4085 (Gt) | | 105,000 |
| *63% hard coal* | *2565 (Gt)* | *32.5E-9 EJ/tonne* | *83,500* |
| *37% lignite coal* | *1520 (Gt)* | *14.0E-9 EJ/tonne* | *21,500* |
| Oil | 5070 (Gb) | | 31,000 |
| *Conventional oil* | *2615 (Gb)* | *6.1E-9 EJ/barrel* | *16,000* |
| *Unconventional oil* | *2455 (Gb)* | *6.1E-9 EJ/barrel* | *15,000* |
| Gas | 675 (Tcm) | | 27,000 |
| *Conventional gas* | *375 (Tcm)* | *40 EJ/Tcm* | *15,000* |
| *Unconventional gas* | *300 (Tcm)* | *40 EJ/Tcm* | *12,000* |
| Total fossil fuels | | | 163,000 |
| Uranium | | | 14,500 |
| **Total nonrenewable energy** | | | 177,500 |

$^*$URR values expressed in EJ have been rounded up to the nearest 500.

## Calibration

Figure 7.3 shows that the global historical patterns of nonrenewable and energy productions (7.3ab), technological change (7.3c), and GWP (7.3d) are acceptably reproduced by the model. The model calibration is particularly good for past global efficiency of primary-to-useful exergy conversion and for past GWP. Regarding nonrenewable energy, the model is not able to start with a zero production in 1750, so it slightly overestimates the historical global trend up to 1950. Concerning the global production of renewable energy (for which data uncertainty is higher than for nonrenewable energy), we have not been able to reproduce the nearly stagnant trend between 1750 and 1910 as the model is only capable of producing continuously increasing dynamics for this variable.

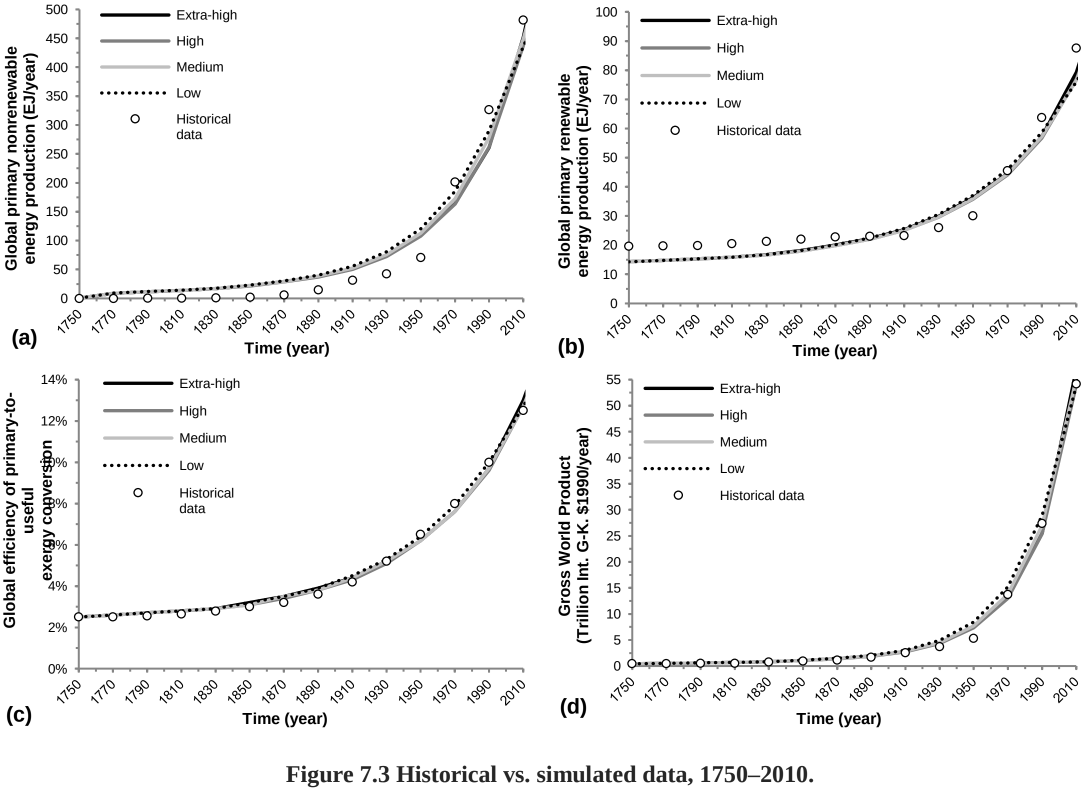


**Figure 7.3 Historical vs. simulated data, 1750–2010.**
**(a) primary nonrenewable energy production, (b) primary renewable energy production, (c) global efficiency of primary-to-useful exergy conversion, (d) gross world product.**

### *7.2.3 Results of Prospective Simulations*

Energy productions, technological change, and GWP

When the model is simulated up to 2250, differences between scenarios clearly appear. Obviously, this is visible in Figure 7.4c where the simulated values of the global efficiency of primary-to-useful exergy conversion are presented for the four scenarios. As formalized in the model, the technological change dynamics directly influence the nonrenewable and renewable energy production paths respectively presented in Figure 7.4a and 7.4b for the four scenarios. Concerning nonrenewable energy, the higher the final technological level the higher the value of the production peak, and possibly the higher the time of that peak (2050 for the *Low* and *Medium* scenarios, 2070 for the *High* and *Extra-high* scenarios). Regarding the renewable energy production, its final value is obviously higher if the final technological level is higher. As can be seen in Figure 7.4d where the GWP is expressed on a log scale for convenience, the energy supply dynamics has a great impact on the economic production. The final *level* of renewable energy production primarily determines the final GWP *level*, but more interestingly the combined dynamics of the nonrenewable and renewable energy productions, i.e. the time path of the energy transition, determine the more or less smoothed course of the GWP. More precisely, if the nonrenewable energy peak is too high compared to the final combination of renewable production and technological level (as in the *Low* and *Medium* scenarios), the GWP can peak and then decrease before stabilizing (the log scale of Figure 7.4d hides this important result of the model, which is more visible in the Figure 7.7c1c2 of Section 7.3.3).

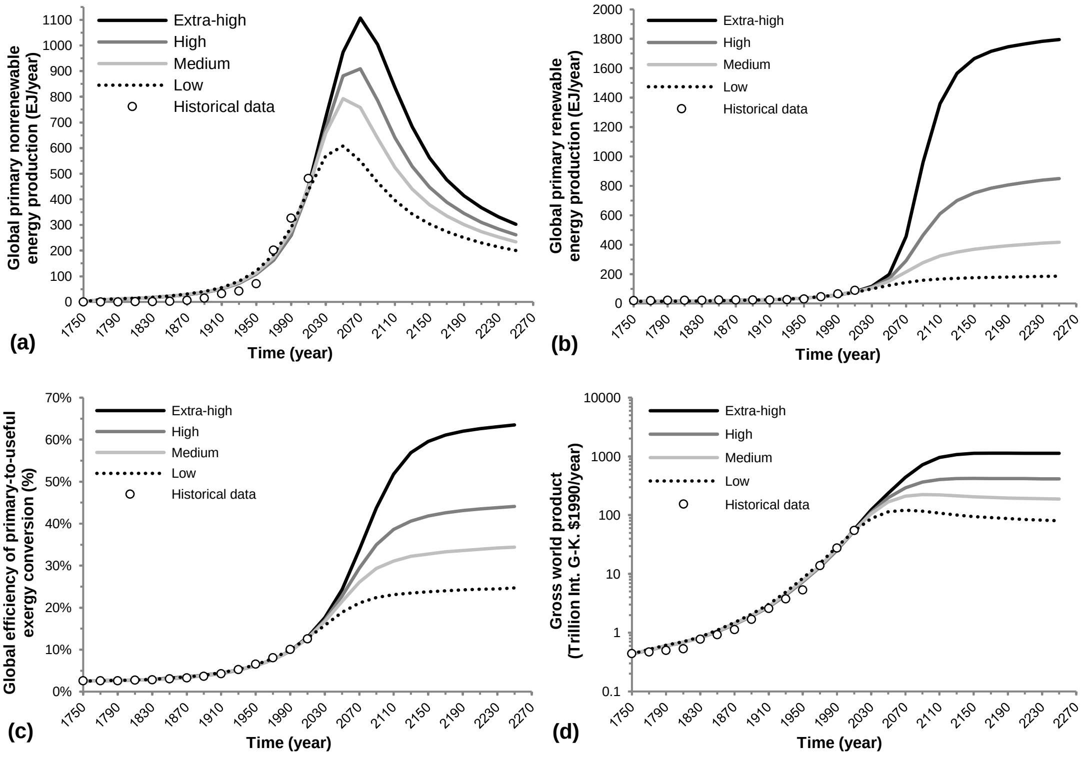


**Figure 7.4 Historical vs. simulated data, 1750–2250.**
**(a) primary nonrenewable energy production, (b) primary renewable energy production, (c) global efficiency of primary-to-useful exergy conversion, (d) gross world product.**

It is important to be clear here: the negative GWP patterns (overshoot before degrowth) of the *Low* and *Medium* scenarios *do not* arise *solely* because their final technological levels $\overline{A}$ (respectively at 0.25 and 0.35) are too low in absolute terms. Rather, the negative impact of the energy transition on economic growth is due in our model to the final value of the technological level *and* the way this variable influences the production cost of the two energy forms. Recall that this link between the technological level $A_t$ of the final good sector and the production costs $D_t(A_t, \phi_{NRE,t})$ and $B_t(A_t)$ (of the nonrenewable and renewable energy sectors respectively) was established to ensure a technological consistency across all sectors of the model. No matter what one would like to change in the parameter settings of the model, the negative impact of the energy transition on economic growth observed for the *Low* and *Medium* scenarios is unavoidable. In Section 7.3 we discuss different strategies that can help smoothing the GWP dynamics in scenarios that initially present the *Low* scenario settings.

Before turning to this section, it is worth analyzing the dynamics of the EROI of the nonrenewable and renewable energy productions since, apart from Fagnart & Germain (2014), our model is the first to introduce these crucial variables in a neoclassical framework.

EROI of energy sector

When computing the EROIs of nonrenewable and renewable energy production, we found quite strange results in light of the EROI literature. Raising these issues is important in order to indicate the features of our model that should deserve particular attention and be improved in future research. First, within a given scenario and a given time period, EROIs of nonrenewable and renewable energy productions have exactly the same value, i.e. $EROI_{NRE,t} = EROI_{RE,t} = EROI_t$. This outcome comes from two particular features of our model: (i) NRE and RE productions are perfect substitutes since they are sold at the same price, (ii) both productions have the same level of self-consumption since we have assumed $\chi_{NRE} = \chi_{RE}$. The first hypothesis is a modeling choice (and we think that including two different prices would not be as simple as one might think because the model would need further complexification in order to remain closed), the second hypothesis is due to the absence of reliable data and therefore the option to choose otherwise.

Second, as shown in Figure 7.5, the EROIs of the economy have relatively low and restricted values (always between 4.1 and 5.8). These low values of the simulated EROIs might surprise people accustomed to the EROI literature. Indeed, generally accepted orders of magnitude of (primary energy) EROI are around 10 for traditional biomass energy (woodfuel, crop residues), 1–2 for modern biofuels, 4–20 for modern renewables (wind, solar, etc.), around 10–30 for conventional oil, 40–60 for gas, and 50–100 for coal (cf. Section 4.2.1). These estimates generally include direct energy consumption in the form of final energy (electricity, liquid fuels, etc.) and indirect energy embodied in *physical* capital. One should not forget that in our model, the quantity of capital *services* $Z_t$ and $G_t$ formally represent labor activated effective capital services or, in other words, the aggregated output of physical capital, routine labor, and human capital.[1] Hence, in our model the denominator of the EROI not only includes

---

[1] The fact that capital services $Z_t$, $G_t$ and $H_t$ represent far more than just physical capital also logically translate in the values of the saving rate $S_t$ varying between 0.5 and 0.8 for all scenarios, which is indeed pretty high compared to current real global saving rates of about 0.22–0.24 (World Bank 2016c).

the energy embodied in physical capital formation but also the energy necessary to sustain labor (i.e. to provide, at a minimum, food and shelter to workers), and to support their skills development. As a consequence, the resulting EROIs of the model represent full lifecycle energy ratios of primary energy production and are thus necessarily quite low compared to conventional values found in the literature that do not take into account such an extended input boundary.

Third, the simulated EROIs have clear U-shapes over the entire time frame, whereas the EROI theoretical dynamic developed in Section 4.1.3 suggests just the opposite. We interpret those U-shapes as the mark of the technological influence in the final good sector. Once the technological level $A_t$ takes off, producing the final output good $Y_t$ requires less and less primary energy, in other words, the energy embodied in capital services decreases. Hence, even if the capital intensiveness of both energy-producing sectors increases due to decreasing marginal returns, their EROI increases after reaching a minimum because the energy embodied in each unit of capital decreases. Logically, when the technological level $A_t$ approaches its asymptotic maximal value, the EROI stabilizes. In order to correct this unrealistic feature of our model, it would be necessary to add another sector and to make a clear distinction between the production of an intermediary capital product and the final good product.

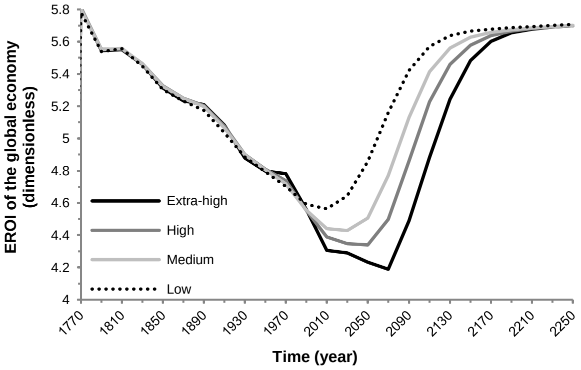


**Figure 7.5 EROI of the global economy, 1770–2250.**

Now that the model results have been analyzed, we can turn our attention to the strategies to avoid the unanticipated nonrenewable energy peak and associated renewable energy supply delay, which cause an overshoot and then degrowth of the economic production in the *Low* and *Medium* scenarios. At first approximation, this energy lock-in which generates this unfortunate GWP dynamics can be thought of as a failure of the price system to incentivize renewable energy production.

## 7.3 DISCUSSION ON THE IMPLEMENTATION OF A CARBON PRICE

We consider that the final technological level $\bar{A}$ of the economy (which is the most important parameter determining the dynamics of the model) can hardly be changed endogenously by a given policy action. This asymptotic value $\bar{A}$ cannot be known a priori but only a posteriori once thermodynamic limits are reached for all the different energy-using devices of the economy (including the ones that have yet to be invented). But the intuition we want to test is that even if this parameter is primordial for determining the *ultimate state* of the economic system, there must be ways to change the *trajectory* that leads to this deterministic end. This point is especially important if this path is believed to generate welfare losses because of the (imposed and not chosen) economy degrowth as in the *Low* and *Medium* scenarios. In other words, the policy actions that must be investigated are the ones that help avoid as much as possible the lock-in phenomenon described previously that is characteristic of both the real world and our model: the tendency of the economic system to stay accustomed to fossil fuels without anticipating their inevitable supply peak and decline and the need associated to increase renewable energy production. Starting from a *Low* scenario setting, the strategy we propose to avoid its adverse outcome (GWP peak followed by a degrowth phase) is to implement a tax on the nonrenewable energy production and to use the income revenue from this tax to direct the energy transition dynamics and smooth its negative impact on the GWP. Such a tax could be indexed to the polluting potential of the fossil energy and more precisely to its greenhouse gases (GHG) content (abstracting from the fact that the nonrenewable resource of the model also contain some GHG-free uranium energy).

Hence, in our model the price that we will exogenously impose on the NRE production could clearly be seen as a *carbon price/tax*. It is important to understand that the income from the carbon pricing can be used in three different ways that can be combined in various proportions to generate many different policy mixes. The annual income from the carbon tax could be used to: (i) subsidize the general R&D sector of the economy in order to accelerate the convergence of $A_t$ towards $\overline{A}$; (ii) subsidize the R&D that is specific to the RE sector in order to accelerate the decrease of $B_t$ towards $\underline{B}$; (iii) subsidize a direct increase in the capital investment $G_t$ in the RE sector. In the rest of this section we will first present the different equation changes resulting from the implementation of the carbon price. Then, the specific mathematical formalization of the uses of the carbon tax income will be successively presented. Finally, we will propose four policy mix scenarios and compare the results of their simulations.

### *7.3.1 Common Equation Changes*

Let us define $q_t$ as the unitary carbon price at period *t* (i.e. the carbon price per unit of pollution, hence expressed in \$/tCO$_2$eq, or B\$/GtCO$_2$eq in order to be consistent with the previous sections). This carbon price is zero prior to the time period $t_{q_t\,start}$ in which it is implemented, and it evolves towards the maximum unitary carbon tax value $\overline{q}$ at exogenous speed $\rho$ following a sigmoid increasing form (Figure 7.6). The maximum growth rate of the unitary carbon tax occurs when $t = t_{q_t\,start} + t_{q_t\,lag}$, so finally

$$q_t = \frac{\overline{q}}{1 + \exp\left(-\rho\left(t - t_{q_t\,start} - t_{q_t\,lag}\right)\right)}, \qquad \forall\, t \geq t_{q_t\,start.} \tag{7.30}$$

Since the NRE producer has to pay the price $q_t$ for every unit of pollution (B\$/GtCO$_2$eq), he has to pay the amount $q_t\kappa$ per unit of nonrenewable energy produced (B\$/EJ), with $\kappa$

representing the GHG emission factor of nonrenewable energy (expressed in $GtCO_2eq/EJ$). Hence, we deduce that the carbon tax income $Q_t$ is defined by

$$Q_t = R_t q_t \kappa, \ \ \forall\, t \geq t_{q_t\, start.} \tag{7.31}$$

Implementing the carbon price also logically changes the equations relative to the NRE producer behavior. More precisely, the implementation of the carbon price leads to the replacement of equations (7.5) and (7.9) with (7.32) and (7.33).

$$\max_{R_t, Z_t} \Pi_t = (1-\chi_{NRE})(p_t - q_t\kappa)R_t - vZ_t\,, \qquad \forall\, t \in \{0,\dots,T_e\} \tag{7.32}$$

$$R_t = \left[\frac{(1-\chi_{NRE})(p_t - q_t\kappa)\theta}{D_t^{\frac{1}{\theta}} v}\right]^{\frac{\theta}{1-\theta}}, \qquad \forall\, t \in \{0,\dots,T_e\} \tag{7.33}$$

In addition to the equation changes that concern the NRE producer previously presented, some equation changes will also be specific to each way of using the carbon tax income $Q_t$.

### *7.3.2 Specific Equation Changes*

Since we have potentially three simultaneous ways to use the income carbon tax, each option represents a share $\beta_i$, $i \in [1,2,3]$ of the total carbon tax income $Q_{t-1}$,with $\beta_1 + \beta_2 + \beta_3 = 1$.

Option (i): carbon price income used to subsidize the general R&D sector

A first option is to allocate the carbon tax income to the general R&D sector in order to increase the growth rate $\xi_t$ of the technological level. Doing so has a direct effect on the GWP with (7.14), and an indirect effect through the impact on the nonrenewable and renewable energy production dynamics with (7.18), (7.22), and (7.24). In order to formalize the use of the income from the carbon tax to subsidize the general R&D sector, we have to replace equation (7.21) defining the speed of convergence $\xi_t$ of the technological level with by the following equation (7.34),

$$\xi_t = \begin{cases} \sigma S_0 + \beta_1 \varepsilon_1 Q_{t-1} & for\ t = 0 \\ \sigma\left[\left(\frac{2}{N+1}\right)S_{t-1} + \left(1 - \frac{2}{N+1}\right)\xi_{t-1}\right] + \beta_1 \varepsilon_1 Q_{t-1}, & \forall\, t \in \{1,\dots,T\}. \end{cases} \tag{7.34}$$

Where $\varepsilon_1$measures the efficiency with which the general R&D sector uses the carbon tax income to produce innovations that materialize in the form of $\xi_t$ increases. The functional form given in (7.34) insures that the higher the carbon tax income of period $t-1$ and the higher the share $\beta_1$ of this income dedicated to the general R&D sector, the faster the technological level will converge towards its upper bound $\overline{A}$.

Option (ii): carbon price income used to subsidize the specific R&D of the RE sector

A second way to use the income from the carbon tax is to allocate it to the R&D that is specifically dedicated to the renewable energy sector. Doing so should affect the rate of degrowth of the unitary capital cost of RE production $B_t$ towards its lower bound $\underline{B}$. An appropriate way to formalize this is to replace (7.24) with the following (7.35),

$$B_t(A_t) = \overline{B} - \frac{\overline{B} - \underline{B}}{1 + \exp(-(\tau + \beta_2 \varepsilon_2 Q_{t-1})(A_t - A_{\Delta B_t max})}, \qquad \forall\, t \in \{0, \dots, T\}. \qquad (7.35)$$

Where $\varepsilon_2$ measures the efficiency with which the specific R&D of the RE sector uses the carbon tax income to produce innovations that materialize in the form of RE production cost decreases. The functional form given in (7.35) insures that the higher the carbon tax income of period $t - 1$ and the higher the share $\beta_2$ of this income dedicated to the specific R&D of the RE sector, the faster the unitary capital cost of RE production will converge towards its lower limit $\underline{B}$.

Option (iii): carbon price income used as a direct capital investment in the RE sector

The third option for using the income from the carbon tax consists of a direct subsidy to the renewable energy sector in order to increase the amount of available energy-capturing capital. This should be seen as the capacity of the RE producer to install an additional amount of physical capital and hire workers thanks to a subsidy from the carbon tax income of the previous period. To formalize this effect we propose to replace (7.11) with the following (7.36),

$$G_t = (F_t B_t)^{\frac{1}{\gamma}} + \beta_3 \varepsilon_3 Q_{t-1}, \quad \forall\, t \in \{0, \dots, T\}. \qquad (7.36)$$

Where $\varepsilon_3$ measures the efficiency with which the RE sector uses the subsidy that is received in the previous period to build new capital and hire additional workers in the RE sector. The functional form given in (7.36) insures that the higher the carbon tax income of period $t - 1$ and the higher the share $\beta_3$, the higher the additional renewable energy produced in period $t$.

### *7.3.3 Results of Simulations*

Defining the policy mixes scenarios and the carbon price profiles

Among the infinity of possibilities, we define four different policy mix scenarios characterized by their relative parameters $\beta_i$, $i \in [1,2,3]$:

- *General R&D* scenario: the totality of the carbon tax income is allocated to the general R&D sector, so $\beta_1 = 1$, $\beta_2 = 0$, and $\beta_3 = 0$.
- *One third each* scenario: the income from the carbon tax is split equally between the three ways of revenue recycling, so $\beta_1 = 1/3$, $\beta_2 = 1/3$, and $\beta_3 = 1/3$.
- *50/50 RE R&D/investment* scenario: the carbon tax income is split equally between the specific R&D of the RE sector and the direct capital investment in

the RE sector, there is no additional subsidy to the general R&D sector, so $\beta_1 = 0$, $\beta_2 = 0.5$, and $\beta_3 = 0.5$.

- *30/70 RE R&D/investment* scenario: 30% of the income from the carbon tax goes to the specific R&D of the RE sector, and 70% is used as a direct capital investment in the RE sector. In this scenario also there is no additional subsidy to the general R&D sector, so $\beta_1 = 0$, $\beta_2 = 0.3$, and $\beta_3 = 0.7$.

We have chosen to test two exogenous carbon price profiles called $q$ and $q'$. They are defined in Table 7.5 by their respective parameters $\overline{q}$, $\rho$, $t_{q_t\,start}$, and $t_{q_t\,lag}$ and shown in Figure 7.6. As previously mentioned we make the hypothesis that all new scenarios in which we implement the carbon price start with the parameter settings of the *Low* scenario. The value of parameter $\kappa$ representing the GHG emission factor of nonrenewable energy (expressed in GtCO$_2$eq/EJ) is set to 0.085 GtCO$_2$eq/EJ. It is the average value found when dividing the historical global GHG emissions from fossil fuels estimated by Boden et al. (2013) by the historical nonrenewable energy production presented in Table 7.1. We consider that $\varepsilon_3$ has the same value as $\lambda$ since both parameters represent productivities of the transformation of investments goods into productive capital, and that there is no apparent reason to think that transformation productivities should differ from one sector of the economy to another. Hence, $\varepsilon_3 = \lambda = 7.25$. On the other hand, since we have no clear way to estimate parameters $\varepsilon_1$ and $\varepsilon_2$, we have arbitrarily chosen the same value of 0.0002 for both parameters, which we found when performing the simulations.

**Table 7.5 Values for parameters defining the two possible carbon taxes *q* and *q'*.**

| Parameter | Definition (unit) | Value for carbon tax $q$ | Value for carbon tax $q'$ |
|---|---|---|---|
| $\overline{q}$ | Maximum level of the carbon tax (Int. G-K. \$1990/tCO$_2$eq) | 400 | 400 |
| $\rho$ | Exogenous growth rate of the carbon tax (dmnl) | 1.4 | 1.0 |
| $t_{q_t\,start}$ | Time period for implementing the carbon tax (time period) | 13 | 13 |
| $t_{q_t\,lag}$ | Time lag to obtain the maximum rate of growth of the carbon tax after its implementation time (time period) | 3 | 5 |

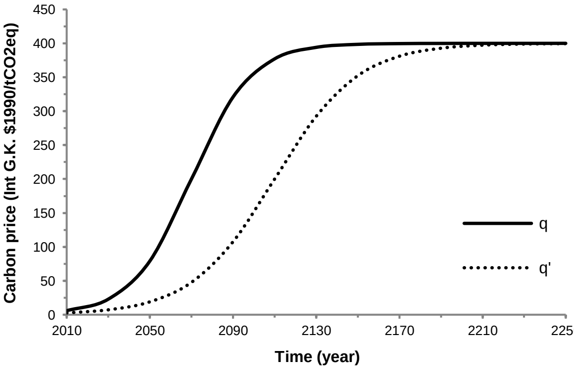


**Figure 7.6 Profiles of the two possible carbon prices *q* and *q'*.**

*Simulation results of General R&D, One third each, 50/50 RE R&D/investment, and 30/70 RE R&D/investment scenarios with carbon prices q and q'*

In Figure 7.7, we compare the nonrenewable and renewable energy productions, and the GWP of the baseline *Low* scenario with the four scenarios that include the carbon tax *q* (left side) or *q'* (right side). Simulations of carbon price scenarios deliver four results. (i) The desired smoothing dynamics of the GWP is only obtained with the *50/50* or *30/70* scenarios in which the carbon tax income is allocated to the specific R&D of the renewable energy sector and to direct capital investment in renewable energy technologies. (ii) The GWP smoothing is higher with the more initially stringent carbon price *q* than with *q'*. (iii) The *General R&D* scenario, and to a lesser extent the *One third each* scenario, leads to a worse situation than the original *Low* scenario, in which the overshoot and degrowth phases of the GWP are accentuated. In these scenarios accelerating the technological change of the final sector exacerbates the nonrenewable energy lock-in of the economy. (iv) This harmful effect of technological and energy resource lock-in is lower if the less stringent carbon price *q'* is chosen.

These results support the criticisms made by Weyant (2011) about the price fundamentalism advanced by Nordhaus (2011). Pricing the externality is not enough, and indeed additional incentives directed specifically to the renewable sector are needed to overcome its market failures, as modeled in the *50/50* and *30/70 RE R&D/investment* scenarios. Of course, further refinements of the model would be needed to correctly define the best policy option for which we do not have an optimization criterion in the current modeling state. Moreover, we have only tested scenarios in which the relative allocation shares $\beta_i$ of the carbon tax revenue remain constant during the entire simulation time, which is of course not the case in the real economy. Nevertheless, implementing the carbon tax in our model was interesting to see that it seems to represent an adequate strategy (among others surely) to attenuate, at least partially, the unfortunate future outcomes suggested by the *Low* scenario.[1]

Our model supports the idea that both the quantity of net exergy supplied by energy-producing sectors to the energy-dissipative economy and the ability of the economic system to use this exergy are key elements of economic growth. To our knowledge, we are the first to develop a simple theoretical model that can be calibrated on global historical data and correctly reproduce long-term global historical trends for nonrenewable and renewable primary energy supply, aggregate technological change, and GWP. This is mainly because, unlike similar approaches, we have ensured that our theoretical model respects some of the many fundamental biophysical limits of the real world. These are formalized in the functional forms that we have established for the capital requirements of nonrenewable and renewable energy productions, and in the technological level of the economy formally defined as the aggregate efficiency of primary-to-useful exergy conversion.

---

[1] Implementing the same smoothing strategy in the *Medium* scenario leads to the same conclusions.

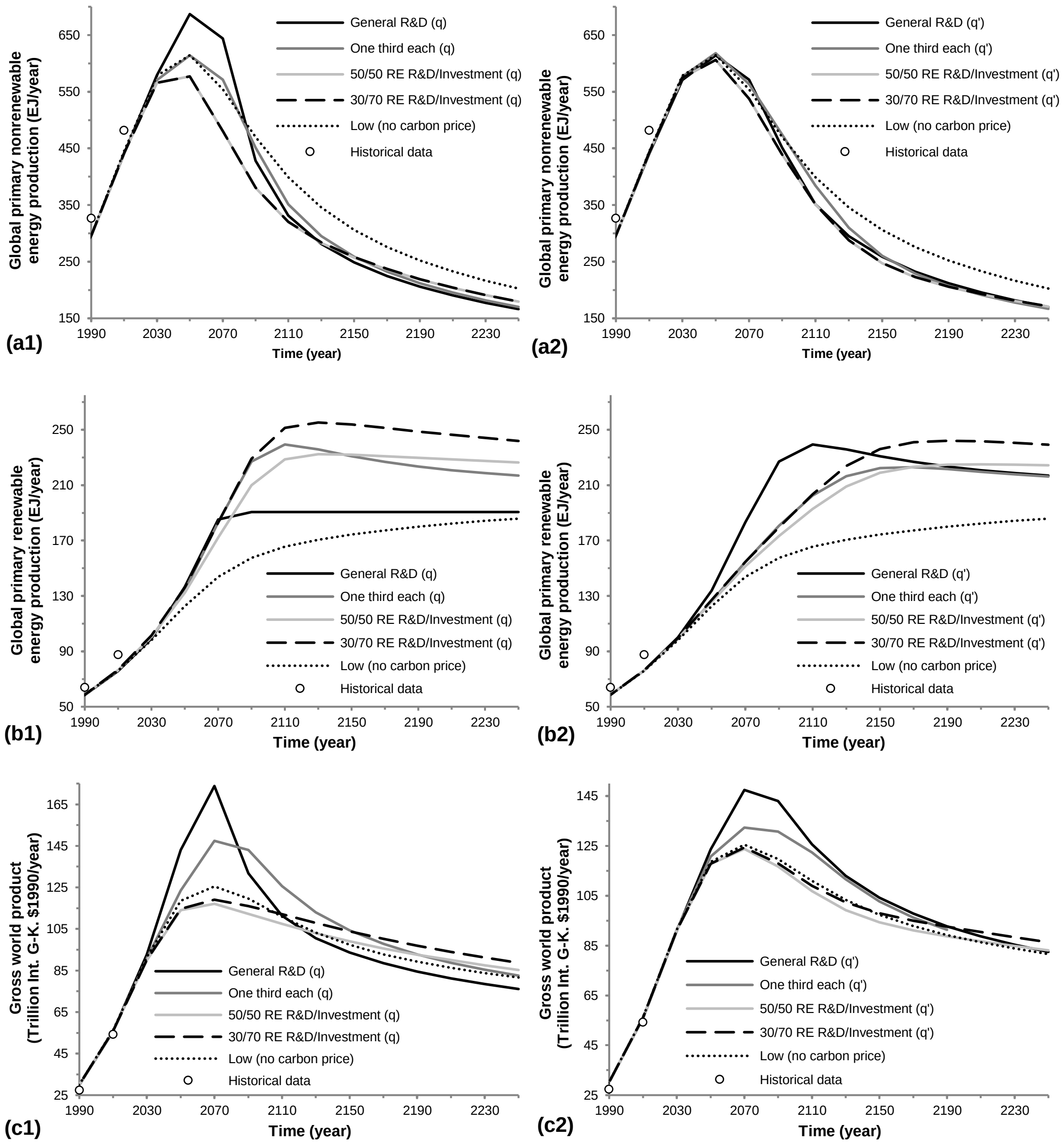


Figure 7.7 Simulation outputs of carbon price scenarios.

Another conclusion of the model presented in this seventh chapter is that for an economy in which energy-producing and energy-consuming sectors are technologically consistent, and in the absence of any correction of the price system, the final efficiency of primary-to-useful exergy conversion of the economy must be high enough (above 0.35) to ensure a smooth future transition from nonrenewable to renewable energy that does not negatively impact economic growth. In our model, the economy cannot avoid a temporary energy lock-in (unanticipated nonrenewable energy peak occurring at a low level of renewable energy production) when this requirement for future technological level is not attained. In such circumstances the energy transition from nonrenewable to renewable energy induces an overshoot and then degrowth of the economic product. Such a lock-in behavior of the economic system can be (at least partially) avoided through the implementation of a carbon price, which also has the benefit of decreasing GHG emissions from fossil-fuels use and hence mitigates climate change. Therefore, implementing a carbon price on nonrenewable energy production and recycling its revenue could help in the choice of the best development path that, at minimum, should consist in a smooth energy transition that does not negatively impact economic development. However, in its current formulation our model cannot be used to define endogenously the optimal time path of the carbon price, nor the optimal time path allocation of the carbon tax revenue among the different recycling uses. This would require us to add some micro-foundations in order to explain how producers and consumers receive adequate incentives to change their behavior.

# CONCLUSION

> *"Clearly, a civilization coping with falling net energy returns will have to make many adjustments. But this reality in itself may not be intolerably restrictive. […] Civilization's course is not preordained but remains open to our choices."*
>
> Vaclav Smil

This thesis is the first step in a broader research project aimed at elaborating a unified theory of economic growth capable of explaining past, current, and future economic growth. The major challenge of such theory is to explain why approximately two hundred years ago some privileged regions (Western Europe and North America) underwent an Industrial Revolution that launched them on a path of relatively sustained high growth rates compared to the previous millennia during which all regions of the world were trapped in a state of Malthusian near-stagnation.

## USEFUL EXERGY CONSUMPTION AS THE FUNDAMENTAL CAUSE OF GROWTH

### The four facts of very long-term economic growth

In order to answer this question, Chapter 1 focused on the description of the four main facts of very long-term economic growth. The transition from stagnation to growth is visible in the growth rates of key variables such as per capita income, population density, fertility, and levels of industrialization and urbanization. The differential timing of the take-off from stagnation to growth among regions of the world and the associated variations in the timing of their demographic transitions led to the phenomenon called the Great Divergence. The initial take-off of England from the Malthusian Epoch was associated with the Industrial Revolution that started there in 1750–60, and then spread to Western Europe and the Western Offshoots (USA, Canada, Australia, and New-Zealand) during the first part of the nineteenth century. The economic take-off of Latin America and West Asia took place toward the beginning of the twentieth century, whereas East Asia and Africa's economic take-offs were further delayed well into the twentieth century. Moreover, the human adventure seems (self-)organized in hierarchized-nested adaptive cycles that define both its structure and its dynamic functioning (Civilizational, Secular, Generational, Kondratieff, Kuznets, Juglar, and Kitchin cycles).

Finally, it is clear that the improvement and diffusion of new technologies, such as general purpose technologies, are closely connected with energy consumption.

Deep-rooted and proximate causes of economic growth

This first chapter gave also a detailed description of the different so-called *ultimate* causes of growth that should preferably be termed *deep-rooted* causes. Several biogeographical factors undeniably had a deep-rooted influence on the timing of the Agricultural and Industrial Revolutions. Those biogeographical factors include favourable climatic conditions, size and orientation of major continental axes, length of the coastline with respect to the mainland size, and mostly the favourable location of primary exergy flows (biomass, then water and wind) and stocks (coal, then oil and gas). Cultural and institutional attributes are interlinked in an intractable endogeneity and seem to be the consequences more than the causes of economic growth and development. Finally, it is possible that some historical events (colonization, silver trade between the Americas and Europe and onward to China from 1500–1800) generated temporary constraints that might have prevented or delayed the economic take-off of several countries.

In order to have a more complete description of the economic growth process, Chapter 2 concentrated on the study of mainstream economic theories that focus on the state of sustained growth attained by industrialized countries. Despite a considerable literature, the proximate causes of economic growth of these theoretical models are always the same. They consist in the accumulation of physical and human capital, and the improvement in the efficiency of the economy to grow output from these inputs. This last variable is called technological change by mainstream economists and corresponds to a catch-all aggregation of many different features of the economic system that, for most of them, should in fact be partially regarded as consequences, or as facilitating factors, of economic growth: the division and organization of labor, the broader organization and efficiency of markets, the improved skills of laborers, the contribution of information and communication technologies, but also the beneficial effects of inclusive institutions (which, for example, protect private property rights and consequently incentivize innovation and R&D). The proximate factors on which conventional economic theory focuses cannot be fundamental causes of growth, for the simple reason that "at some level (and exaggerating somewhat) to say that a country is poor because it has insufficient physical capital, human capital, and inefficient technology is like saying that a person is poor because he or she does not have money" (Acemoglu 2009, p.106).

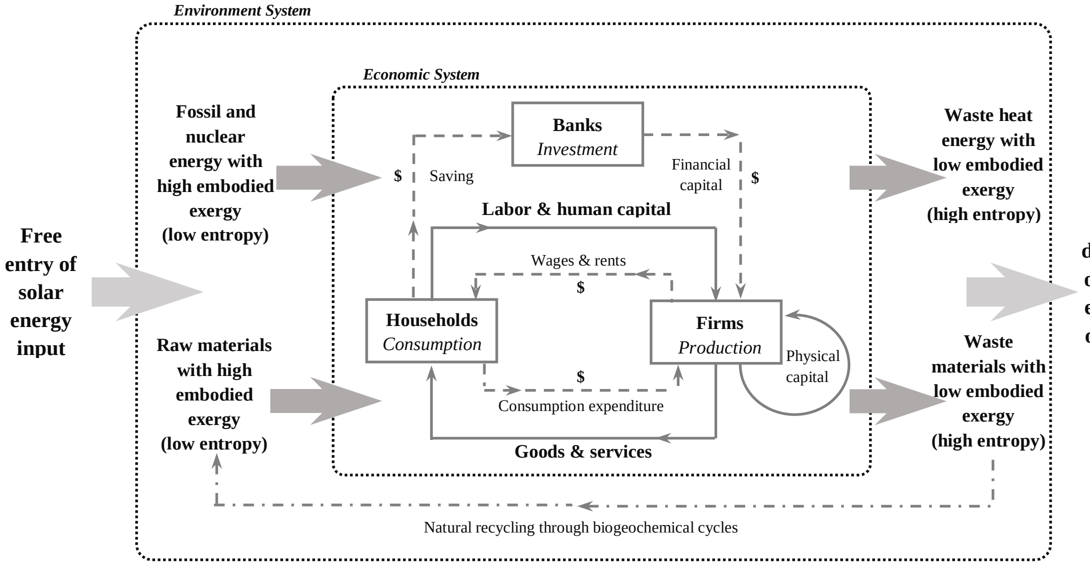
If the proximate causes of growth (physical and human capital accumulation, technological change) remain the same in all the different approaches presented in this chapter, the way they are interlinked in Unified Growth Theory is more ambitious. Indeed, compared to the more conventional settings of neoclassical economics that focus on the Modern Growth Era, UGT is a far more powerful framework in which to investigate the transition from stagnation to sustained growth. UGT has undoubtedly brought a new perspective to the process of economic growth, but its core mechanisms are clearly questionable. UGT will be a more promising avenue for correctly understanding long-term economic growth once mechanisms developed on a stronger empirical basis are incorporated into its analytical framework.

This extensive review of all mainstream economic growth theories was indispensable to see that they fail to properly explain the growth process and that they are completely disconnected from any biophysical reality.

### Exergy, entropy, and economic growth

Chapter 3 started by analyzing the misguided reasons for overlooking natural resources, and in particular energy, in mainstream economics. Then, essential concepts such as exergy and entropy were described to explain that the first and second laws of thermodynamics always apply to the economic system and shape its functioning. Applying these fundamental laws to the economic system demonstrates that mainstream economists see the economy as a perpetual motion machine of the first kind, that is, a machine that performs work though endless cycles without any input of energy. Hence, the mainstream conception of the economic system is a conceptual artifact that can absolutely not exist in the real world. Energy makes up a small share of total production costs not because it is less important than capital or labor as a production factor, but rather because the biosphere and geosphere generate the physical work that we use abundantly and free-of-charge.

As put by Atkins (2010, p. 22), if the first law of thermodynamics was found to be false, "wealth–and untold benefits to humanity–would accrue to an untold extent". The second law of thermodynamics is also essential to understanding that in reality the economic system is an open system (in the thermodynamic sense) that extracts and converts low entropy matter-energy into even lower entropy products (and services) and gives off high entropy wastes that are freely discarded in the environment (see Graphical synthesis 1). The unspontaneous decrease in entropy associated with the increasing order of matter from raw to refined materials in the forms of goods is only possible because an even higher amount of entropy production is associated with the degradation of exergy extracted from the environment.



**Graphical synthesis 1. The economic system as an exergy-degrading real machine.**

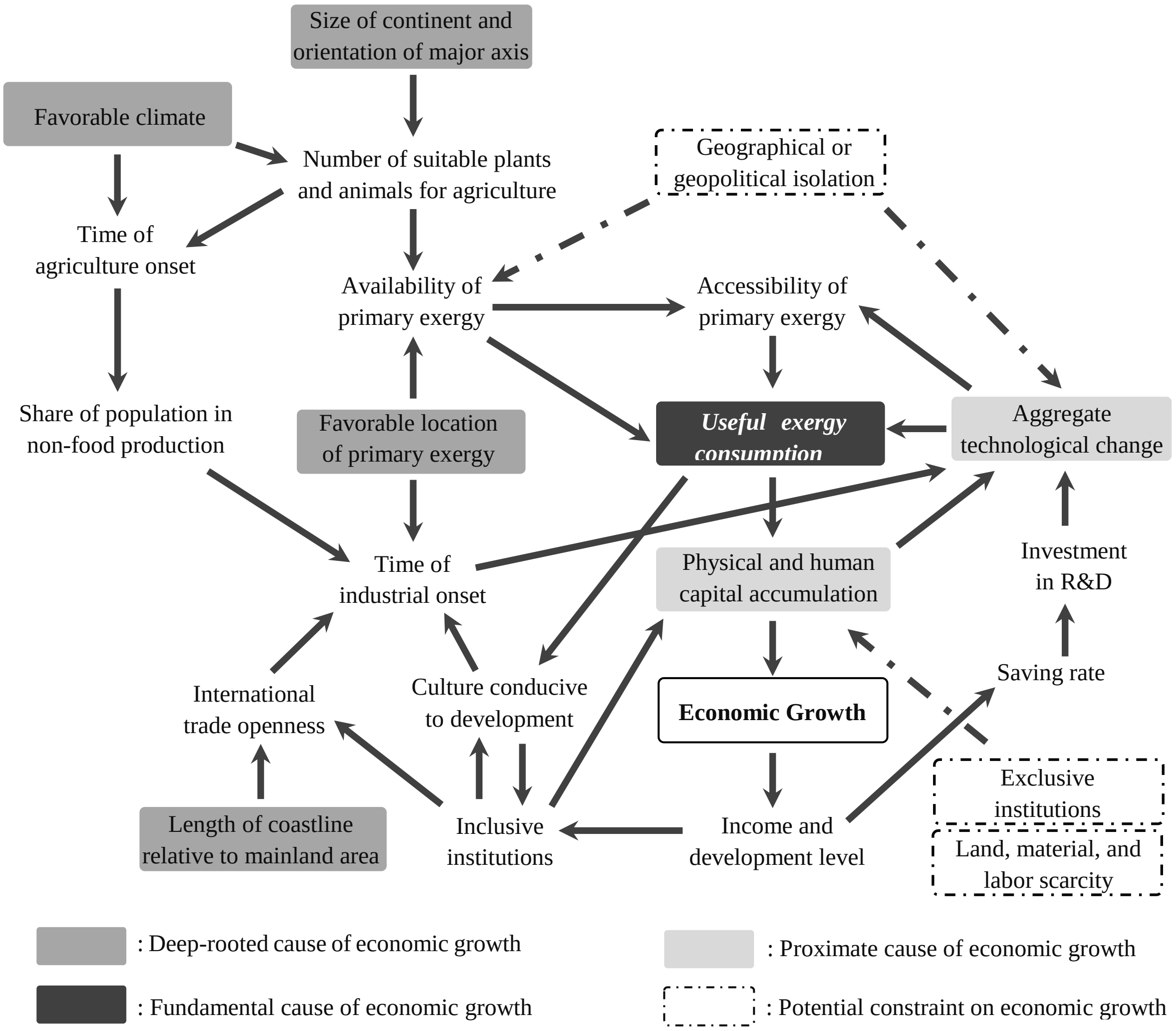
This biophysical approach shows that only useful exergy consumption can be considered as the fundamental cause of economic growth. The capacity to extract primary exergy from the environment and the ability to convert it with increasing efficiency into useful exergy services (in the form of light, heat, electricity, and mechanical power) is clearly the principal mechanism explaining economic growth. With such an approach, technological change can be precisely defined as gains in the aggregate efficiency of primary-to-useful exergy conversion. It is understandable , then, that structural limits of materials will define the ultimate limit of the aggregate technological level of the economy, and this ultimate limit might be closer than expected. In particular, the stagnation of the aggregate efficiency of primary-to-useful exergy conversion in industrialized countries since the 1970s seems to be an important cause of the slowdown in economic growth endured by those same countries for the last forty years. Moreover, a biophysical approach to the economic system shows that the so-called dematerialization of the economy, or decoupling of GDP, is a pure illusion that has no logical reality.

### Useful exergy consumption and societal development

As Sieferle (1997) put it, “universal history can be subdivided into three parts. Each part is characterized by a certain energy system [foraging, farming, fossil fuel burning]. This energy system establishes the general framework, within which the structures of society, economy, and culture form. Thus, energy is not just one factor acting among many. Rather, it is possible, in principle, to determine the formal basic structures of a society from the pertaining energetic system conditions”. Regarding the causal relation between systems of energy capture and value systems adopted by societies, it is important to stress that “it is not that individuals are caused to adopt values by their society’s mode of energy capture. Rather, over the course of long stretches of history, and as a result of innumerable social experiments by inventive humans, the societies that are best organized to exploit available modes of energy capture–by their social structures, economic and political institutions, culture and values–will tend to prevail over and displace other societies that are less well organized. Social forms and the associated values that are ill adapted to human survival and comfort, given available technologies, will give way to more effective institutions and values” (Stephen Macedo in the introduction to Morris 2015, p. XIX).

In agrarian societies, economic growth depends on the capacity of people (that are often coerced as slaves or serfs) to extract increasing solar exergy in the form of food, fodder, motion (from water and wind exergy flows, which are derived from solar exergy), and woodfuel (here again indirect solar exergy), and on the ability to transform those primary exergy resources into useful exergy in the form of light, heat, and mechanical power. On the other hand, during the last two centuries, animal and human labor has been gradually replaced by fossil-exergy-activated machines which have driven down the cost of goods and services (in terms of the number of working hours required to buy such products) and have consequently increased demand and production. In current developed countries, this long-term substitution seems to have been the dominant driver of economic growth since the Industrial Revolution (Ayres & Warr 2009, p.168). More recently, transistors powered by electricity have started to further reduce biological limitations as they assist the human brain in processing and storing huge quantities of information. Hence, in modern industrialized societies, it is “exergy that drives the machines in mines and on drilling sites, in power stations, factories and office buildings, on rails, road and farms, in the air, and on the sea. In short, it activates the wealth-creating production process of industrial economies" (Kümmel 2011, p.37).

To sum up, the so-called *causes* of growth generally emphasized by standard economics, such as inclusive institutions (property rights protection, democratic expression), conducive cultural traits (scientific and hard-work spirits), Smithian features (market size, competition, variety of products, division of labor) and others (international trade openness), should more appropriately be regarded as facilitating factors and even consequences of growth. The fundamental cause of economic growth is useful exergy consumption, which is the combined capacity to extract primary exergy from the environment and the ability to transform it into useful services in the form of light, heat, electricity, and mechanical power. The differences between fundamental, deep-rooted, and proximate causes of growth and its consequences are summarized in the Graphical Synthesis 2.



**Graphical synthesis 2. Fundamental, deep-rooted, and proximate causes of growth.**

The first part of this thesis (Chapters 1 to 3) highlights the adequacy of the biophysical/thermo economics approach for understanding the economic growth phenomenon. If conventional theories are unable to correctly explain the four long-term facts of economic growth (transition from stagnation to growth, Great Divergence, energy consumption-technological change interdependence, and hierarchized-nested adaptive cycles dynamics), this thesis shows that, at least for the first three of them, the role played by energy in the economic system is primordial. Higher energy availability and accessibility are predominant in explaining that the onset of the Industrial Revolution occurred in Britain and not elsewhere. The local energy availability and accessibility, and the magnitude and time differences in the spread of technologies that enable an increase in the aggregate primary-to-useful exergy conversion efficiency have largely defined the direction of the Great Divergence. As a consequence, the future economic growth of countries will depend essentially on (i) the continued increase in the aggregate efficiency of primary-to-useful exergy conversion, and/or (ii) the continued increase in the extraction of available primary exergy resources. The former point has already been discussed but the later needs to be addressed in terms of net energy (exergy).

## NET EXERGY CONSTRAINT IN THE COMING CENTURY

### Energy-return-on-investment as a measure of accessibility

The availability of an energy resource is given by its level of ultimately recoverable resource (URR) in the case of a non-renewable stock and its technical potential (TP) in the case of a renewable flow. However, the ease of extracting an energy resource, that is to say, its accessibility, is given by its energy-return-on-investment, or EROI. The EROI is a crucial indicator of development because all societies need energy resources that deliver more energy than is invested to use them. Furthermore, it seems logical to think that all different types of societies have a notional minimum EROI required to sustain their level of development. Chapter 4 presents in detail the static (meaning for a given representative year) calculation methodology of the EROI of a given energy system, the different controversies surrounding such a calculation, and hence the limits of this concept. Studies have shown that fossil energy resources to which modern economies have become accustomed and on which they are dependent do not generate as much net energy as before. Indeed, all studies estimating EROI time-series of fossil fuels have so far reached the same result: declining trends in recent decades with maximum EROI already passed. A price-based methodology developed in this same chapter put this issue in a new perspective. It showed that maximum EROI has indeed already been reached at global level for oil and gas production, but for coal, net energy gains are still to be expected thanks to forthcoming technological improvements. On the other hand, recent studies show that unconventional fossil fuels do not generate as much net energy as conventional fossil energy used to do. Most importantly, renewable technologies, which policy makers and many experts see as humanity’s future, present EROIs that are (currently) lower than past and current fossil fuel EROIs, especially when the intermittent nature of these renewable energy resources is taken into account. Of course, there is great scope for improvements in these immature technologies, but for them too, the *First Best Principle* that consists in the use of the best resources first before turning towards lower quality resources applies. Hence, all economies will eventually head towards a future in which ever more energy is invested in the energy-extraction sub-system of the economy, making net energy delivered to society less readily available.

EROI and qualitative depletion of metals

Chapter 5 shed some light on the close relationship between the energy and metal sectors from the EROI perspective. First, we supported the position of Rankin (2011) by estimating that 10% of global primary energy production is consumed by the metal sector. Then, we showed that the energy consumption of the metal sector has increased faster than the rest of the economy since 1973. Supported by previous studies, we made the fair assumption that this apparent increasing energy requirement of the metal extraction sector is mainly due to decreasing ore grade. The decline in quality of ores is a natural process that occurs across different level (deposit, nation, world) and implies that more energy is needed to extract a given quantity of metal. Because renewable technologies have higher metal intensities than conventional means of electricity production, the question of the sustainability of a transition consisting in a shift toward renewables is legitimate, especially because those energy systems have lower EROIs than fossil fuels. Logically, we decided to estimate how the energy requirement associated with metal extraction could impact the EROIs of different electricity producing technologies.

A first analysis consisted in calculating the sensitivity of the EROIs of renewable and nuclear technologies assuming different levels of ore grade degradation for a specific metal. We explained the kind of results that it is possible to obtain through examples of three metal (copper, nickel, and chromium), although this kind of sensitivity calculation could have been performed for any metal used in a given technology. Each technology displays a specific sensitivity to a particular metal that can be measured through the methodology we have developed. In a second step, we adapted our methodology in order to calculate the sensitivity of the EROIs of the same technologies to a similar depletion of all rare metals. This exercise was useful to see that energy requirements associated with metal extraction could have a significant impact on the capacity of these “green” technologies to deliver net energy to society. Of course, the question of the speed of degradation of the average ore grade of a given metal remains unanswered. This evolution will be different for each metal but will ultimately have a negative impact on the EROI of renewable technologies.

In the context of a transition toward renewables that are more metal-intensive than fossil energy systems, all other things being equal, the increasing energy requirement of the metal sector due to metal ore grade degradation will further increase the demand for renewable energy. Moreover, the intermittency of these technologies implies the need to expand and reinforce the transmission grid and storage capacities, which will generate an even greater demand for metals. As a consequence, in the perspective of a transition toward renewable technologies, a potential vicious circle could develop between the energy and metal sectors. It is currently impossible to say whether such an unpleasant situation would effectively arise but this chapter has started a quantitative exploration of this issue.

Energy expenditures, economic growth and the minimum EROI of society

Chapter 6 presented estimations of the level of energy expenditure, i.e. the amount of GDP diverted to obtain energy, from 1850 to 2012 for the US and the global economy, and from 1300 to 2008 for the UK. Results indicate that the level of energy expenditure in the economy, seems to play a limit-to-growth role since as long as it has remained above 6–8% of GDP, high economic growth rates have never occurred in the last several centuries for the US, the UK, and even the global economy. More precisely, periods of high or suddenly increasing energy expenditure levels are associated with low economic growth rates: for instance from 1850 to 1945 (very high energy expenditure levels), from 1975 to 1976 (surge), and from 1981 to 1983 (surge). On the contrary, periods of low and decreasing energy expenditure are associated with high and increasing economic growth rates: for instance from 1945 to 1973, and in the early 2000s.

Furthermore, we were able to show that in order to have a positive growth rate, from a statistical point of view, the US economy cannot afford to allocate more than 11% of its GDP to primary energy expenditure (in the absence of other major limits of a geographical, geopolitical, or institutional nature). This means that considering its current energy intensity, the US economy needs to have at least a societal $EROI_{min}$ of approximately 11:1 (that conversely corresponds to a maximum tolerable average price of energy of twice the current level) in order to present positive rates of growth.

Finally, over the more restricted period 1960–2010 for which we have continuous year-to-year data for the US, we performed several Granger causality tests that consistently show a one way negative causality running from the level of energy expenditure (as a fraction of GDP) to economic growth.

Economic growth and energy transitions

Chapter 7 builds a bridge between the endogenous economic growth theory, the biophysical economics perspective, and the past and future transitions between renewable and nonrenewable energy forms that economies have had and will have to accomplish. The model supports the evidence that historical productions of renewable and nonrenewable energy have greatly influenced past economic growth. Indeed, from an initial almost-renewable-only supply regime, the model reproduces the increasing reliance on nonrenewable energy that has allowed the global economy to leave the state of economic near-stagnation that characterized the largest part of its history.

The model supports the idea that both the quantity of net exergy supplied by energy-producing sectors to the energy-dissipative economy and the ability of the economic system to use this exergy are key elements of economic growth. Unlike similar approaches, the theoretical model respects some of the many fundamental biophysical limits of the real world. These are formalized in the functional forms that we have established for the capital requirements of nonrenewable and renewable energy productions, and in the technological level of the economy formally defined as the aggregate efficiency of primary-to-useful exergy conversion.

The main result of the model presented in this seventh chapter is that for a global economy in which energy-producing and energy-consuming sectors are technologically consistent, and in the absence of any correction of the price system, the final efficiency of primary-to-useful exergy conversion of the economy must be high enough (above 0.35) to ensure a smooth future transition from nonrenewable to renewable energy that does not negatively impact economic growth. In our model, the global economy cannot avoid a temporary energy lock-in (unanticipated nonrenewable energy peak occurring at a low level of renewable energy production) when this requirement for future technological level is not attained. In such circumstances the energy transition from nonrenewable to renewable energy induces an overshoot and then degrowth of the economic product. Such a lock-in behavior of the economic system can be (at least partially) avoided through the implementation of a carbon price, which also has the benefit of decreasing GHG emissions from fossil-fuels use and hence mitigates climate change.

The second part of this thesis (Chapters 4 to 7) suggests that maintaining a future high net energy supply is likely to become increasingly difficult given the past evolution of fossil fuel EROIs and considering the current low EROIs of renewable energy-producing technologies towards which industrial societies are supposed to make a transition. There are of course significant opportunities for maintaining a high societal EROI or adapting to decreasing EROIs. But from a systemic point of view, industrialized societies seem not to be designed to run with low-density energy resources that come with low EROIs. Until proven otherwise, high economic growth is only possible if high-density energy resources infuse the economic system and allow physical and human capital accumulation, the establishment of inclusive institutions, higher material standards of living, higher qualitative leisure, and in summary higher welfare for people.

## RESEARCH PERSPECTIVES

It is essential to build a unified theory of economic growth for two reasons. First, the understanding of the contemporary growth process will always remain incomplete and fragile if growth theory cannot reflect within a single framework the various qualitative aspects of societal development over the entire human adventure. For so long as the economic take-off encountered by some privileged countries two hundred years ago remains a mystery, confidence in modern economic growth can only be fragile. Second, a comprehensive understanding of the obstacles faced by less-developed countries in reaching a state of sustained economic growth can only be achieved if the factors that prompted the transition of the currently developed economies to a state of sustained economic growth can be identified and their implications modified to allow for the differences in the growth structure of less-developed economies in an interdependent world.

The research begun in this thesis will require further work in order to develop a unified theory of economic growth that respects the biophysical constraints of the real world. As stated in the introduction, a multidisciplinary approach is needed to make a success of such a project. A truly unified theory of economic growth will only be achieved once the role of exergy consumption will be correctly taken into account and accurately linked with other determinants of the growth process, namely the deep-rooted and proximate causes of economic growth.

# APPENDICES

## APPENDIX A: GROWTH ACCOUNTING

The following growth accounting framework is a mixed adaptation of Barro & Sala-i-Martin (2004) and Acemoglu (2009). It starts from a standard production function, which we can write as

$$Y = F(A, K, L). \tag{A.1}$$

Where $Y$ is GDP, $A$ is the level of technology, $K$ is the capital stock, and $L$ is the quantity of labor. Capital and labor can be disaggregated among types or qualities as in Jorgenson & Griliches (1967). The production function makes clear that GDP can grow only if there is growth in productive inputs, including the level of technology. The growth rate of output can be partitioned into components associated with factor accumulation and technological progress. Taking logarithms of equation (A.1) and derivatives with respect to time we get

$$\frac{\dot{Y}}{Y} = \frac{F_A A}{Y}\frac{\dot{A}}{A} + \frac{F_K K}{Y}\frac{\dot{K}}{K} + \frac{F_L}{Y}\frac{\dot{L}}{L}. \tag{A.2}$$

Where $F_A$, $F_K$, and $F_L$ are the marginal products of $A$, $K$, and $L$, respectively defined by

$$F_A = \frac{\partial F(A, K, L)}{\partial A}, \qquad F_K = \frac{\partial F(A, K, L)}{\partial K}, \qquad and\ F_L = \frac{\partial F(A, K, L)}{\partial L}. \tag{A.3}$$

These definition mean that $\varepsilon_K \equiv F_K K/Y$ and $\varepsilon_L \equiv F_L L/Y$ are the respective elasticities of output with respect to capital and labor. And usually, the growth rates of output, capital and labor are respectively denoted as $g \equiv \dot{Y}/Y$, $g_K \equiv \dot{K}/K$, and $g_L \equiv \dot{L}/L$. The contribution of technology is noted $x \equiv \frac{F_A A}{Y}\frac{\dot{A}}{A}$, hence

$$g = x + \varepsilon_K g_K + \varepsilon_L g_L. \tag{A.4}$$

Equation (A.4) is no more than an identity saying that the growth rate of GDP can be decomposed into the weighted growth rates of the three inputs: technology, capital, and labor. And in particular, the weights are given by the relative contributions of each of the factors to GDP. These contributions, in turn, are the social marginal products times the amount of input divided by GDP, i.e. the output elasticities of inputs. This formulation includes Hicks-neutral and labor-augmenting technological progress as special cases. If the technology factor appears in a Hicks-neutral way, so that $F(A,K,L) = A \cdot \tilde{F}(K,L)$, then $F_A A = Y$ and $x = \dot{A}/A$. If the technology factor appears in a labor-augmenting form, so that $F(A,K,L) = \tilde{F}(K,AL)$ then $F_A A = F_L L$ and $x = \varepsilon_L \dot{A}/A$.

To estimate $x$ empirically, we need to know the growth rates of $Y$, $K$, and $L$, and the social marginal products $F_K$ and $F_L$. The former are can be measured empirically (although not without difficulty) but the later would typically not be measurable directly. In practice, researchers assume that the social marginal products can be measured by observed factor prices. If the factors are paid to their social marginal products, so that $F_K = R$ the rental price of capital, and $F_L = w$ the wage rate, then $F_K K = RK$ is the total amount of capital rents paid in the economy (the rent bill), and $F_L L = wL$ is the total amount of wages paid in the economy (the wage bill). Hence, $F_K K/Y = RK/Y$ is the fraction of GDP used to rent capital, a fraction known as the *capital share,* which we denote by $s_K$. Similarly, $F_L L/Y = wL/Y$ is the fraction of GDP used to pay wages, a fraction known as the *labor share,* which we denote by $s_L$. Using these notations, the estimation $\hat{x}$ of the rate of technological progress can be rewritten as

$$\hat{x} = g - s_K g_K - s_L g_L. \tag{A.5}$$

The value $\hat{x}$ is often described as an estimate of total factor productivity (TFP) growth. This formulation was first presented by Solow (1957), so the value $\hat{x}$ is also sometimes called the *Solow residual*. Since the method just described relies on the growth rates of the quantities of inputs, the label *primal* is sometimes attached to *TFP growth* or to *Solow residual*. This labeling distinguishes this approach from the *dual* price-based method which is not described here.

Early studies, such as Solow (1957) and Denison (1967), found very large residuals. In other words, a substantial fraction of the growth rate of aggregate output was not accounted for by the growth rates of measured inputs, and, consequently, a substantial role was assigned to technological progress. Jorgenson & Griliches (1967) showed that a substantial fraction of the Solow residual could be explained by changes in the quality of inputs. For example, improvements in the quality of the labor force reflect increases in average years of schooling and better health. For given quantities of capital and worker hours, improvements in the quality of labor raise output. But if labor input is measured only by worker hours, the unmeasured quality improvements show up as TFP growth. Unmeasured improvements in the quality of capital have similar effects.

To take improvements in the quality of labor into account, worker hours can be disaggregated into many different categories based on schooling, experience, gender, and so on (see Jorgenson et al. 1987). Each category is weighed in accordance with its observed average wage rate, the usual proxy for the marginal product of labor. For example, if persons with college education have higher wage rates (and are presumably more productive) than persons with high school education, then an extra worker with a college education accounts for more output expansion than would an extra worker with a high school education.

In this approach, the overall labor input is the weighted sum over all categories, where the weights are the relative wage rates. For a given total of worker hours, the quality of the labor force improves—and, hence, the measured labor input increases—if workers shift toward the categories that pay higher wage rates. For example, if the fraction of the labor force that is college educated increases and the fraction with no schooling declines, then the total labor input rises even if the aggregate amount of worker hours does not change.

The allowance for quality change in the capital stock also requires a disaggregation into many components. The aggregate measure of capital input is the weighted sum over all types, where the weights are the relative rental rates. To compute the rental rates, the usual assumption is that all investments yield the same rate of return.

# APPENDIX B: GLOBAL PRIMARY ENERGY PRODUCTIONS, 1800-2014.

Several studies deal with historical primary energy productions but (to my best knowledge) it is impossible to find a unique publication providing coherent time series for all the different primary energy forms that have been used at global scale during the last two hundred years. I have summarized in Table B1 the different time series I have consolidated and to which I refer in Section 1.2.3 and Section 7.2.1 of the present thesis and in some of my peer-review publications.

Regarding traditional biomass energy (woodfuel and crop residues), I averaged data from Fernandes et al. (2007) and Smil (2010). Primary fossil fuels time series were retrieved from The Shift Project (2015), which is built on the original work of Etemad & Luciani (1991) for the 1900–1980 time period and EIA (2014) for 1981–2010. From 1800 to 1900, I have completed the different fossil fuel time series with the original 5-years interval data of Etemad & Luciani (1991) and filled the gaps using linear interpolation. The online data portal of The Shift Project (2015) was also used to retrieve productions of nuclear and renewable electricity, i.e. hydro, wind, solar, geothermal, wastes, ocean (wave, tidal, OTEC) electricity, and for modern biofuels (ethanol and biodiesel). Concerning nuclear and renewable electricity values, a correction was brought to the original time series to correct a commonly agreed, yet absolutely not scientifically backed-up, convention: when expressed in primary equivalent terms, renewable electricity productions are usually expressed in *raw electricity* terms, whereas nuclear electricity is expressed artificially boosted in *heat equivalent* terms. Indeed, generally when one speaks about a primary energy mix, a three-fold factor is systematically applied to nuclear raw electricity production estimates to take into account that (uranium) atoms fission first generate heat used to boil water into steam which kinetic energy is converted into electricity with an overall average efficiency of 33%. I have no problem with such a convention to express nuclear energy in primary term, but I strongly support that electricity from so-called renewable technologies should suffer the same kind of conventional arithmetic to express their production estimates in primary terms. Hence, in the same way that a 33% efficiency is commonly assumed for nuclear primary-to-final energy conversion, I have applied equivalent factors to renewable electricity producing technologies. Based on Kreith & Goswami (2007) and Zarrouk & Moon (2014), I have slightly changed the primary-to-final energy conversion efficiency provided by the EIA (2012, p.345): 33% for nuclear, 85% for hydropower, 25% for wind power, 15% for solar, 12% for geothermal power, 33% for biomass/wastes, and 50% for wave/tidal plants. Human-food energy was estimated by hypothesizing an average daily intake of 2500 kcal/capita which was multiplied to a year-to-year global population estimate based on the original data of the United Nations (1999, p.5; 2015). Draft animal-food energy (i.e. fodder) and traditional water/wind energy use (through waterwheel, windmill, and sail ships), were estimated using the following backward induction process: (i) the shares of these energy forms in the global supply mix have been arbitrarily chosen at different time step based on Kander et al. (2013) and basic linear interpolations was used to produce continuous times series for these relative shares; (ii) a counterfactual total global energy consumption including food, fodder and traditional water/wind uses is computed using the previously determined relative shares; (iii) multiplying fodder and traditional water/wind relative shares to the counterfactual total delivers year-to-year energy consumption estimates of fodder and traditional water/wind energy.

**Table B1. Global primary energy production, 1800–2014.**

| Time | Coal (EJ/yr) | Oil (EJ/yr) | Gas (EJ/yr) | Nuclear (EJ/yr) | Hydro (EJ/yr) | Wind (EJ/yr) | Solar (EJ/yr) | Geothermal (EJ/yr) | Wastes (EJ/yr) | Wave/ Tidal (EJ/yr) | Modern biofuels (EJ/yr) | Woodfuel/ Crop residues (EJ/yr) | Food (EJ/yr) | Fodder (EJ/yr) | Trad. water/wind (EJ/yr) |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1800 | 0.310 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 20.000 | 3.873 | 2.684 | 0.268 |
| 1801 | 0.333 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 20.041 | 3.875 | 2.677 | 0.267 |
| 1802 | 0.355 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 20.082 | 3.876 | 2.669 | 0.266 |
| 1803 | 0.378 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 20.121 | 3.878 | 2.661 | 0.264 |
| 1804 | 0.400 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 20.161 | 3.880 | 2.653 | 0.263 |
| 1805 | 0.423 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 20.199 | 3.882 | 2.645 | 0.261 |
| 1806 | 0.445 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 20.237 | 3.884 | 2.637 | 0.260 |
| 1807 | 0.468 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 20.275 | 3.886 | 2.629 | 0.258 |
| 1808 | 0.490 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 20.312 | 3.888 | 2.620 | 0.257 |
| 1809 | 0.513 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 20.349 | 3.891 | 2.612 | 0.255 |
| 1810 | 0.536 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 20.385 | 3.893 | 2.603 | 0.253 |
| 1811 | 0.558 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 20.420 | 3.895 | 2.594 | 0.252 |
| 1812 | 0.581 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 20.456 | 3.898 | 2.585 | 0.250 |
| 1813 | 0.603 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 20.492 | 3.900 | 2.577 | 0.249 |
| 1814 | 0.626 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 20.529 | 3.903 | 2.568 | 0.247 |
| 1815 | 0.648 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 20.567 | 3.906 | 2.559 | 0.246 |
| 1816 | 0.671 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 20.604 | 3.909 | 2.550 | 0.244 |
| 1817 | 0.693 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 20.643 | 3.912 | 2.542 | 0.242 |
| 1818 | 0.716 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 20.681 | 3.915 | 2.533 | 0.241 |
| 1819 | 0.738 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 20.721 | 3.918 | 2.524 | 0.239 |
| 1820 | 0.761 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 20.760 | 3.921 | 2.515 | 0.238 |
| 1821 | 0.783 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 20.800 | 3.924 | 2.507 | 0.237 |
| 1822 | 0.806 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 20.840 | 3.928 | 2.498 | 0.236 |
| 1823 | 0.828 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 20.880 | 3.931 | 2.489 | 0.235 |
| 1824 | 0.851 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 20.920 | 3.935 | 2.480 | 0.234 |

| Time | Coal (EJ/yr) | Oil (EJ/yr) | Gas (EJ/yr) | Nuclear (EJ/yr) | Hydro (EJ/yr) | Wind (EJ/yr) | Solar (EJ/yr) | Geothermal (EJ/yr) | Wastes (EJ/yr) | Wave/ Tidal (EJ/yr) | Modern biofuels (EJ/yr) | Woodfuel/ Crop residues (EJ/yr) | Food (EJ/yr) | Fodder (EJ/yr) | Trad. water/wind (EJ/yr) |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1825 | 0.873 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 20.960 | 3.939 | 2.471 | 0.233 |
| 1826 | 0.896 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 21.000 | 3.943 | 2.462 | 0.232 |
| 1827 | 0.918 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 21.039 | 3.947 | 2.453 | 0.231 |
| 1828 | 0.941 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 21.079 | 3.951 | 2.444 | 0.230 |
| 1829 | 0.963 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 21.119 | 3.955 | 2.434 | 0.229 |
| 1830 | 0.986 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 21.158 | 3.960 | 2.425 | 0.228 |
| 1831 | 1.021 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 21.198 | 3.964 | 2.417 | 0.227 |
| 1832 | 1.057 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 21.236 | 3.969 | 2.409 | 0.227 |
| 1833 | 1.092 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 21.273 | 3.974 | 2.400 | 0.226 |
| 1834 | 1.128 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 21.309 | 3.979 | 2.392 | 0.225 |
| 1835 | 1.164 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 21.344 | 3.985 | 2.383 | 0.224 |
| 1836 | 1.199 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 21.378 | 3.990 | 2.374 | 0.223 |
| 1837 | 1.235 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 21.411 | 3.996 | 2.365 | 0.222 |
| 1838 | 1.270 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 21.442 | 4.001 | 2.356 | 0.221 |
| 1839 | 1.306 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 21.473 | 4.007 | 2.346 | 0.220 |
| 1840 | 1.341 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 21.502 | 4.014 | 2.337 | 0.219 |
| 1841 | 1.426 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 21.531 | 4.020 | 2.331 | 0.216 |
| 1842 | 1.510 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 21.557 | 4.027 | 2.325 | 0.213 |
| 1843 | 1.594 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 21.581 | 4.034 | 2.319 | 0.210 |
| 1844 | 1.679 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 21.603 | 4.041 | 2.312 | 0.208 |
| 1845 | 1.763 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 21.622 | 4.048 | 2.305 | 0.205 |
| 1846 | 1.847 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 21.640 | 4.055 | 2.298 | 0.201 |
| 1847 | 1.932 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 21.655 | 4.063 | 2.291 | 0.198 |
| 1848 | 2.016 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 21.669 | 4.071 | 2.284 | 0.195 |
| 1849 | 2.100 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 21.680 | 4.080 | 2.276 | 0.192 |
| 1850 | 2.185 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 22.054 | 4.088 | 2.297 | 0.191 |
| 1851 | 2.335 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 22.064 | 4.097 | 2.294 | 0.189 |

| Time | Coal (EJ/yr) | Oil (EJ/yr) | Gas (EJ/yr) | Nuclear (EJ/yr) | Hydro (EJ/yr) | Wind (EJ/yr) | Solar (EJ/yr) | Geothermal (EJ/yr) | Wastes (EJ/yr) | Wave/ Tidal (EJ/yr) | Modern biofuels (EJ/yr) | Woodfuel/ Crop residues (EJ/yr) | Food (EJ/yr) | Fodder (EJ/yr) | Trad. water/wind (EJ/yr) |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1852 | 2.484 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 22.075 | 4.106 | 2.291 | 0.186 |
| 1853 | 2.634 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 22.085 | 4.116 | 2.288 | 0.183 |
| 1854 | 2.784 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 22.095 | 4.126 | 2.284 | 0.180 |
| 1855 | 2.933 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 22.106 | 4.136 | 2.280 | 0.177 |
| 1856 | 3.083 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 22.116 | 4.146 | 2.277 | 0.174 |
| 1857 | 3.232 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 22.126 | 4.157 | 2.273 | 0.171 |
| 1858 | 3.382 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 22.136 | 4.168 | 2.268 | 0.168 |
| 1859 | 3.532 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 22.147 | 4.180 | 2.264 | 0.165 |
| 1860 | 3.681 | 0.003 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 22.157 | 4.192 | 2.260 | 0.161 |
| 1861 | 3.909 | 0.006 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 22.217 | 4.204 | 2.265 | 0.159 |
| 1862 | 4.137 | 0.009 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 22.278 | 4.217 | 2.270 | 0.156 |
| 1863 | 4.365 | 0.012 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 22.338 | 4.230 | 2.275 | 0.154 |
| 1864 | 4.593 | 0.015 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 22.398 | 4.243 | 2.279 | 0.151 |
| 1865 | 4.820 | 0.018 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 22.459 | 4.257 | 2.283 | 0.148 |
| 1866 | 5.048 | 0.021 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 22.519 | 4.272 | 2.287 | 0.145 |
| 1867 | 5.276 | 0.024 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 22.579 | 4.287 | 2.290 | 0.142 |
| 1868 | 5.504 | 0.027 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 22.639 | 4.302 | 2.293 | 0.139 |
| 1869 | 5.731 | 0.030 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 22.700 | 4.318 | 2.296 | 0.136 |
| 1870 | 5.959 | 0.033 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 22.760 | 4.335 | 2.301 | 0.133 |
| 1871 | 6.294 | 0.048 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 22.798 | 4.352 | 2.307 | 0.130 |
| 1872 | 6.630 | 0.063 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 22.836 | 4.369 | 2.316 | 0.127 |
| 1873 | 6.965 | 0.080 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 22.873 | 4.387 | 2.324 | 0.124 |
| 1874 | 7.300 | 0.097 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 22.911 | 4.406 | 2.332 | 0.120 |
| 1875 | 7.635 | 0.116 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 22.949 | 4.425 | 2.339 | 0.117 |
| 1876 | 7.971 | 0.135 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 22.987 | 4.445 | 2.347 | 0.114 |
| 1877 | 8.306 | 0.154 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 23.025 | 4.466 | 2.353 | 0.110 |
| 1878 | 8.641 | 0.175 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 23.062 | 4.487 | 2.360 | 0.106 |

| Time | Coal (EJ/yr) | Oil (EJ/yr) | Gas (EJ/yr) | Nuclear (EJ/yr) | Hydro (EJ/yr) | Wind (EJ/yr) | Solar (EJ/yr) | Geothermal (EJ/yr) | Wastes (EJ/yr) | Wave/ Tidal (EJ/yr) | Modern biofuels (EJ/yr) | Woodfuel/ Crop residues (EJ/yr) | Food (EJ/yr) | Fodder (EJ/yr) | Trad. water/wind (EJ/yr) |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1879 | 8.976 | 0.195 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 23.100 | 4.509 | 2.366 | 0.103 |
| 1880 | 9.312 | 0.179 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 23.138 | 4.532 | 2.369 | 0.099 |
| 1881 | 9.782 | 0.204 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 23.117 | 4.555 | 2.372 | 0.098 |
| 1882 | 10.252 | 0.229 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 23.096 | 4.579 | 2.375 | 0.097 |
| 1883 | 10.722 | 0.255 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 23.074 | 4.604 | 2.377 | 0.096 |
| 1884 | 11.192 | 0.280 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 23.053 | 4.630 | 2.379 | 0.095 |
| 1885 | 11.662 | 0.306 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 23.032 | 4.656 | 2.380 | 0.094 |
| 1886 | 12.132 | 0.331 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 23.011 | 4.684 | 2.380 | 0.093 |
| 1887 | 12.603 | 0.357 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 22.990 | 4.712 | 2.379 | 0.092 |
| 1888 | 13.073 | 0.382 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 22.968 | 4.741 | 2.378 | 0.091 |
| 1889 | 13.543 | 0.408 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 22.947 | 4.771 | 2.376 | 0.090 |
| 1890 | 14.013 | 0.433 | 0.238 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 22.926 | 4.803 | 2.387 | 0.090 |
| 1891 | 14.538 | 0.475 | 0.241 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 22.840 | 4.835 | 2.384 | 0.088 |
| 1892 | 15.062 | 0.516 | 0.244 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 22.755 | 4.868 | 2.381 | 0.087 |
| 1893 | 15.587 | 0.557 | 0.246 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 22.669 | 4.902 | 2.377 | 0.086 |
| 1894 | 16.112 | 0.598 | 0.249 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 22.584 | 4.937 | 2.372 | 0.084 |
| 1895 | 16.636 | 0.640 | 0.252 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 22.498 | 4.974 | 2.366 | 0.083 |
| 1896 | 17.161 | 0.681 | 0.255 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 22.412 | 5.011 | 2.360 | 0.081 |
| 1897 | 17.686 | 0.722 | 0.257 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 22.327 | 5.050 | 2.353 | 0.080 |
| 1898 | 18.210 | 0.763 | 0.260 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 22.241 | 5.090 | 2.346 | 0.078 |
| 1899 | 18.735 | 0.805 | 0.263 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 22.156 | 5.131 | 2.338 | 0.077 |
| 1900 | 19.260 | 0.846 | 0.266 | 0.000 | 0.012 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 22.070 | 5.173 | 2.330 | 0.075 |
| 1901 | 19.608 | 0.946 | 0.295 | 0.000 | 0.014 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 22.172 | 5.217 | 2.325 | 0.073 |
| 1902 | 20.082 | 1.027 | 0.323 | 0.000 | 0.015 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 22.274 | 5.262 | 2.324 | 0.072 |
| 1903 | 21.914 | 1.098 | 0.352 | 0.000 | 0.018 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 22.377 | 5.309 | 2.386 | 0.072 |
| 1904 | 22.026 | 1.233 | 0.381 | 0.000 | 0.020 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 22.479 | 5.357 | 2.369 | 0.070 |
| 1905 | 23.746 | 1.213 | 0.410 | 0.000 | 0.023 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 22.581 | 5.406 | 2.416 | 0.070 |

| Time | Coal (EJ/yr) | Oil (EJ/yr) | Gas (EJ/yr) | Nuclear (EJ/yr) | Hydro (EJ/yr) | Wind (EJ/yr) | Solar (EJ/yr) | Geothermal (EJ/yr) | Wastes (EJ/yr) | Wave/ Tidal (EJ/yr) | Modern biofuels (EJ/yr) | Woodfuel/ Crop residues (EJ/yr) | Food (EJ/yr) | Fodder (EJ/yr) | Trad. water/wind (EJ/yr) |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1906 | 25.076 | 1.208 | 0.438 | 0.000 | 0.026 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 22.683 | 5.457 | 2.445 | 0.069 |
| 1907 | 27.738 | 1.494 | 0.467 | 0.000 | 0.030 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 22.785 | 5.510 | 2.542 | 0.070 |
| 1908 | 26.361 | 1.615 | 0.496 | 0.000 | 0.034 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 22.888 | 5.564 | 2.454 | 0.065 |
| 1909 | 27.546 | 1.697 | 0.525 | 0.000 | 0.039 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 22.990 | 5.620 | 2.475 | 0.064 |
| 1910 | 29.024 | 1.845 | 0.567 | 0.000 | 0.043 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 23.092 | 5.678 | 2.509 | 0.063 |
| 1911 | 29.541 | 1.955 | 0.597 | 0.000 | 0.048 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 23.179 | 5.737 | 2.411 | 0.057 |
| 1912 | 30.999 | 1.986 | 0.628 | 0.000 | 0.051 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 23.265 | 5.798 | 2.341 | 0.052 |
| 1913 | 33.215 | 2.190 | 0.658 | 0.000 | 0.058 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 23.352 | 5.861 | 2.299 | 0.047 |
| 1914 | 29.812 | 2.256 | 0.688 | 0.000 | 0.069 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 23.438 | 5.926 | 2.055 | 0.039 |
| 1915 | 29.340 | 2.421 | 0.718 | 0.000 | 0.076 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 23.525 | 5.993 | 1.919 | 0.032 |
| 1916 | 31.472 | 2.597 | 0.748 | 0.000 | 0.085 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 23.611 | 6.062 | 1.860 | 0.027 |
| 1917 | 32.775 | 2.871 | 0.778 | 0.000 | 0.091 | 0.000 | 0.000 | 0.001 | 0.000 | 0.000 | 0.000 | 23.698 | 6.133 | 1.771 | 0.020 |
| 1918 | 32.189 | 2.851 | 0.809 | 0.000 | 0.103 | 0.000 | 0.000 | 0.001 | 0.000 | 0.000 | 0.000 | 23.784 | 6.206 | 1.621 | 0.014 |
| 1919 | 27.301 | 3.246 | 0.839 | 0.000 | 0.058 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 23.871 | 6.281 | 1.385 | 0.006 |
| 1920 | 31.290 | 4.159 | 0.869 | 0.000 | 0.150 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 23.957 | 6.358 | 1.363 | 0.000 |
| 1921 | 26.129 | 4.503 | 0.899 | 0.000 | 0.143 | 0.000 | 0.000 | 0.001 | 0.000 | 0.000 | 0.000 | 24.064 | 6.438 | 1.236 | 0.000 |
| 1922 | 27.578 | 5.026 | 0.960 | 0.000 | 0.161 | 0.000 | 0.000 | 0.001 | 0.000 | 0.000 | 0.000 | 24.170 | 6.520 | 1.247 | 0.000 |
| 1923 | 32.123 | 5.945 | 1.124 | 0.000 | 0.197 | 0.000 | 0.000 | 0.001 | 0.000 | 0.000 | 0.000 | 24.277 | 6.604 | 1.324 | 0.000 |
| 1924 | 31.280 | 5.928 | 1.289 | 0.000 | 0.217 | 0.000 | 0.000 | 0.001 | 0.000 | 0.000 | 0.000 | 24.384 | 6.691 | 1.279 | 0.000 |
| 1925 | 31.359 | 6.215 | 1.454 | 0.000 | 0.251 | 0.000 | 0.000 | 0.001 | 0.000 | 0.000 | 0.000 | 24.491 | 6.780 | 1.257 | 0.000 |
| 1926 | 31.219 | 6.379 | 1.619 | 0.000 | 0.299 | 0.000 | 0.000 | 0.001 | 0.000 | 0.000 | 0.000 | 24.597 | 6.872 | 1.228 | 0.000 |
| 1927 | 33.641 | 7.283 | 1.783 | 0.000 | 0.334 | 0.000 | 0.000 | 0.002 | 0.000 | 0.000 | 0.000 | 24.704 | 6.966 | 1.253 | 0.000 |
| 1928 | 33.044 | 7.649 | 1.948 | 0.000 | 0.376 | 0.000 | 0.000 | 0.002 | 0.000 | 0.000 | 0.000 | 24.811 | 7.063 | 1.218 | 0.000 |
| 1929 | 35.142 | 8.574 | 2.069 | 0.000 | 0.449 | 0.000 | 0.000 | 0.002 | 0.000 | 0.000 | 0.000 | 24.917 | 7.163 | 1.233 | 0.000 |
| 1930 | 32.112 | 8.176 | 2.169 | 0.000 | 0.412 | 0.000 | 0.000 | 0.002 | 0.000 | 0.000 | 0.000 | 25.024 | 7.265 | 1.145 | 0.000 |
| 1931 | 28.320 | 7.911 | 2.272 | 0.000 | 0.441 | 0.000 | 0.000 | 0.002 | 0.000 | 0.000 | 0.000 | 25.101 | 7.371 | 1.051 | 0.000 |
| 1932 | 25.180 | 7.551 | 2.375 | 0.000 | 0.459 | 0.000 | 0.000 | 0.002 | 0.000 | 0.000 | 0.000 | 25.177 | 7.479 | 0.969 | 0.000 |

| Time | Coal (EJ/yr) | Oil (EJ/yr) | Gas (EJ/yr) | Nuclear (EJ/yr) | Hydro (EJ/yr) | Wind (EJ/yr) | Solar (EJ/yr) | Geothermal (EJ/yr) | Wastes (EJ/yr) | Wave/ Tidal (EJ/yr) | Modern biofuels (EJ/yr) | Woodfuel/ Crop residues (EJ/yr) | Food (EJ/yr) | Fodder (EJ/yr) | Trad. water/wind (EJ/yr) |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1933 | 26.362 | 8.245 | 2.478 | 0.000 | 0.491 | 0.000 | 0.000 | 0.002 | 0.000 | 0.000 | 0.000 | 25.254 | 7.590 | 0.964 | 0.000 |
| 1934 | 28.765 | 8.701 | 2.581 | 0.000 | 0.517 | 0.000 | 0.000 | 0.002 | 0.000 | 0.000 | 0.000 | 25.330 | 7.705 | 0.970 | 0.000 |
| 1935 | 30.005 | 9.460 | 2.684 | 0.000 | 0.607 | 0.000 | 0.000 | 0.002 | 0.000 | 0.000 | 0.000 | 25.407 | 7.822 | 0.962 | 0.000 |
| 1936 | 32.960 | 10.302 | 2.787 | 0.000 | 0.645 | 0.000 | 0.000 | 0.003 | 0.000 | 0.000 | 0.000 | 25.483 | 7.943 | 0.973 | 0.000 |
| 1937 | 34.628 | 11.726 | 2.990 | 0.000 | 0.740 | 0.000 | 0.000 | 0.001 | 0.000 | 0.000 | 0.000 | 25.560 | 8.067 | 0.974 | 0.000 |
| 1938 | 32.308 | 11.447 | 3.175 | 0.000 | 0.740 | 0.000 | 0.000 | 0.006 | 0.000 | 0.000 | 0.000 | 25.636 | 8.194 | 0.907 | 0.000 |
| 1939 | 34.127 | 11.926 | 3.403 | 0.000 | 0.773 | 0.000 | 0.000 | 0.015 | 0.000 | 0.000 | 0.000 | 25.713 | 8.324 | 0.895 | 0.000 |
| 1940 | 37.640 | 12.310 | 3.631 | 0.000 | 0.802 | 0.000 | 0.000 | 0.016 | 0.000 | 0.000 | 0.000 | 25.789 | 8.458 | 0.896 | 0.000 |
| 1941 | 39.017 | 12.681 | 3.859 | 0.000 | 0.850 | 0.000 | 0.000 | 0.019 | 0.000 | 0.000 | 0.000 | 25.862 | 8.595 | 0.895 | 0.000 |
| 1942 | 39.686 | 11.936 | 4.087 | 0.000 | 0.875 | 0.000 | 0.000 | 0.027 | 0.000 | 0.000 | 0.000 | 25.934 | 8.736 | 0.876 | 0.000 |
| 1943 | 40.786 | 12.845 | 4.315 | 0.000 | 0.995 | 0.000 | 0.000 | 0.027 | 0.000 | 0.000 | 0.000 | 26.007 | 8.880 | 0.879 | 0.000 |
| 1944 | 39.162 | 14.753 | 4.544 | 0.000 | 0.953 | 0.000 | 0.000 | 0.008 | 0.000 | 0.000 | 0.000 | 26.079 | 9.028 | 0.863 | 0.000 |
| 1945 | 31.423 | 14.736 | 4.772 | 0.000 | 0.938 | 0.000 | 0.000 | 0.003 | 0.000 | 0.000 | 0.000 | 26.152 | 9.179 | 0.777 | 0.000 |
| 1946 | 33.833 | 15.675 | 5.027 | 0.000 | 1.048 | 0.000 | 0.000 | 0.007 | 0.000 | 0.000 | 0.000 | 26.224 | 9.335 | 0.792 | 0.000 |
| 1947 | 38.015 | 17.268 | 5.562 | 0.000 | 1.097 | 0.000 | 0.000 | 0.020 | 0.000 | 0.000 | 0.000 | 26.297 | 9.494 | 0.829 | 0.000 |
| 1948 | 39.279 | 19.544 | 6.288 | 0.000 | 1.209 | 0.000 | 0.000 | 0.026 | 0.000 | 0.000 | 0.000 | 26.369 | 9.657 | 0.847 | 0.000 |
| 1949 | 37.345 | 19.432 | 6.719 | 0.000 | 1.250 | 0.000 | 0.000 | 0.032 | 0.000 | 0.000 | 0.000 | 26.442 | 9.824 | 0.816 | 0.000 |
| 1950 | 40.797 | 21.800 | 7.693 | 0.000 | 1.359 | 0.000 | 0.000 | 0.038 | 0.000 | 0.000 | 0.000 | 26.514 | 9.650 | 0.850 | 0.000 |
| 1951 | 42.981 | 24.722 | 9.127 | 0.000 | 1.538 | 0.000 | 0.000 | 0.048 | 0.000 | 0.000 | 0.000 | 26.908 | 9.828 | 0.886 | 0.000 |
| 1952 | 42.734 | 25.986 | 9.831 | 0.000 | 1.640 | 0.000 | 0.000 | 0.055 | 0.000 | 0.000 | 0.000 | 27.302 | 10.004 | 0.882 | 0.000 |
| 1953 | 42.923 | 27.477 | 10.397 | 0.000 | 1.702 | 0.000 | 0.000 | 0.056 | 0.000 | 0.000 | 0.000 | 27.696 | 10.180 | 0.882 | 0.000 |
| 1954 | 42.370 | 28.766 | 10.906 | 0.000 | 1.806 | 0.000 | 0.000 | 0.056 | 0.000 | 0.000 | 0.000 | 28.090 | 10.359 | 0.875 | 0.000 |
| 1955 | 46.036 | 32.240 | 11.863 | 0.000 | 1.954 | 0.000 | 0.000 | 0.056 | 0.000 | 0.000 | 0.000 | 28.484 | 10.541 | 0.916 | 0.000 |
| 1956 | 48.651 | 35.085 | 12.869 | 0.001 | 2.151 | 0.000 | 0.000 | 0.053 | 0.000 | 0.000 | 0.000 | 28.878 | 10.728 | 0.943 | 0.000 |
| 1957 | 50.018 | 36.975 | 14.000 | 0.005 | 2.339 | 0.000 | 0.000 | 0.054 | 0.000 | 0.000 | 0.000 | 29.272 | 10.920 | 0.955 | 0.000 |
| 1958 | 51.598 | 37.901 | 15.075 | 0.005 | 2.569 | 0.000 | 0.000 | 0.058 | 0.000 | 0.000 | 0.000 | 29.666 | 11.119 | 0.961 | 0.000 |
| 1959 | 53.478 | 40.899 | 16.832 | 0.016 | 2.646 | 0.000 | 0.000 | 0.063 | 0.000 | 0.000 | 0.000 | 30.060 | 11.324 | 0.984 | 0.000 |

| Time | Coal (EJ/yr) | Oil (EJ/yr) | Gas (EJ/yr) | Nuclear (EJ/yr) | Hydro (EJ/yr) | Wind (EJ/yr) | Solar (EJ/yr) | Geothermal (EJ/yr) | Wastes (EJ/yr) | Wave/ Tidal (EJ/yr) | Modern biofuels (EJ/yr) | Woodfuel/ Crop residues (EJ/yr) | Food (EJ/yr) | Fodder (EJ/yr) | Trad. water/wind (EJ/yr) |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1960 | 55.713 | 44.067 | 18.316 | 0.030 | 2.884 | 0.000 | 0.000 | 0.069 | 0.000 | 0.000 | 0.000 | 30.454 | 11.534 | 0.820 | 0.000 |
| 1961 | 52.487 | 46.906 | 19.672 | 0.048 | 3.045 | 0.000 | 0.000 | 0.083 | 0.000 | 0.000 | 0.000 | 30.720 | 11.751 | 0.813 | 0.000 |
| 1962 | 53.753 | 50.881 | 21.509 | 0.071 | 3.206 | 0.000 | 0.000 | 0.088 | 0.000 | 0.000 | 0.000 | 30.985 | 11.975 | 0.836 | 0.000 |
| 1963 | 55.391 | 54.591 | 23.527 | 0.122 | 3.371 | 0.000 | 0.000 | 0.101 | 0.000 | 0.000 | 0.000 | 31.251 | 12.206 | 0.858 | 0.000 |
| 1964 | 57.847 | 58.956 | 25.699 | 0.169 | 3.499 | 0.000 | 0.000 | 0.112 | 0.000 | 0.000 | 0.000 | 31.516 | 12.446 | 0.887 | 0.000 |
| 1965 | 58.996 | 63.252 | 27.536 | 0.269 | 3.859 | 0.000 | 0.000 | 0.119 | 0.000 | 0.000 | 0.000 | 31.782 | 12.697 | 0.907 | 0.000 |
| 1966 | 59.816 | 68.635 | 29.630 | 0.384 | 4.136 | 0.000 | 0.000 | 0.122 | 0.000 | 0.000 | 0.000 | 32.048 | 12.957 | 0.930 | 0.000 |
| 1967 | 57.696 | 73.711 | 31.882 | 0.463 | 4.219 | 0.000 | 0.000 | 0.130 | 0.000 | 0.000 | 0.000 | 32.313 | 13.227 | 0.937 | 0.000 |
| 1968 | 59.313 | 80.494 | 34.875 | 0.575 | 4.418 | 0.000 | 0.000 | 0.131 | 0.000 | 0.000 | 0.000 | 32.579 | 13.505 | 0.970 | 0.000 |
| 1969 | 60.513 | 86.636 | 38.298 | 0.679 | 4.679 | 0.000 | 0.000 | 0.144 | 0.000 | 0.000 | 0.000 | 32.844 | 13.787 | 0.999 | 0.000 |
| 1970 | 62.925 | 95.275 | 42.392 | 0.859 | 4.904 | 0.000 | 0.000 | 0.142 | 0.000 | 0.000 | 0.000 | 33.110 | 14.072 | 1.043 | 0.000 |
| 1971 | 62.355 | 101.073 | 45.324 | 1.183 | 5.144 | 0.000 | 0.000 | 0.139 | 0.000 | 0.000 | 0.000 | 33.408 | 14.360 | 1.057 | 0.000 |
| 1972 | 62.134 | 106.682 | 48.020 | 1.616 | 5.378 | 0.000 | 0.000 | 0.164 | 0.000 | 0.000 | 0.000 | 33.706 | 14.650 | 1.070 | 0.000 |
| 1973 | 63.242 | 116.409 | 50.713 | 2.163 | 5.463 | 0.000 | 0.000 | 0.182 | 0.000 | 0.000 | 0.000 | 34.004 | 14.941 | 1.102 | 0.000 |
| 1974 | 63.795 | 116.813 | 51.844 | 2.786 | 5.996 | 0.000 | 0.000 | 0.207 | 0.000 | 0.000 | 0.000 | 34.302 | 15.231 | 1.090 | 0.000 |
| 1975 | 67.591 | 110.901 | 52.015 | 3.837 | 6.061 | 0.000 | 0.000 | 0.238 | 0.000 | 0.000 | 0.000 | 34.600 | 15.520 | 1.063 | 0.000 |
| 1976 | 69.516 | 120.300 | 54.240 | 4.474 | 6.088 | 0.000 | 0.000 | 0.259 | 0.000 | 0.000 | 0.000 | 34.898 | 15.807 | 1.089 | 0.000 |
| 1977 | 71.625 | 124.639 | 59.375 | 5.510 | 6.234 | 0.000 | 0.000 | 0.267 | 0.000 | 0.000 | 0.000 | 35.196 | 16.093 | 1.107 | 0.000 |
| 1978 | 72.736 | 126.018 | 57.556 | 6.456 | 6.758 | 0.000 | 0.000 | 0.252 | 0.000 | 0.000 | 0.000 | 35.494 | 16.380 | 1.087 | 0.000 |
| 1979 | 77.000 | 130.844 | 62.062 | 6.711 | 7.066 | 0.000 | 0.000 | 0.329 | 0.000 | 0.000 | 0.000 | 35.792 | 16.670 | 1.106 | 0.000 |
| 1980 | 68.136 | 135.811 | 57.774 | 6.860 | 7.297 | 0.000 | 0.000 | 0.397 | 0.187 | 0.003 | 0.079 | 36.090 | 16.966 | 1.054 | 0.000 |
| 1981 | 68.590 | 128.626 | 58.631 | 7.787 | 7.398 | 0.000 | 0.000 | 0.439 | 0.188 | 0.004 | 0.095 | 36.496 | 17.268 | 1.011 | 0.000 |
| 1982 | 71.032 | 123.311 | 58.551 | 8.647 | 7.582 | 0.000 | 0.000 | 0.470 | 0.289 | 0.004 | 0.142 | 36.902 | 17.575 | 0.979 | 0.000 |
| 1983 | 71.162 | 122.950 | 59.205 | 9.866 | 7.928 | 0.000 | 0.000 | 0.533 | 0.300 | 0.004 | 0.201 | 37.308 | 17.889 | 0.957 | 0.000 |
| 1984 | 75.195 | 126.421 | 65.158 | 11.919 | 8.191 | 0.001 | 0.000 | 0.607 | 0.332 | 0.004 | 0.280 | 37.714 | 18.212 | 0.975 | 0.000 |
| 1985 | 77.788 | 125.558 | 67.656 | 14.039 | 8.267 | 0.001 | 0.000 | 0.669 | 0.337 | 0.004 | 0.301 | 38.121 | 18.544 | 0.963 | 0.000 |
| 1986 | 79.701 | 130.605 | 68.926 | 15.133 | 8.436 | 0.002 | 0.000 | 0.746 | 0.365 | 0.004 | 0.281 | 38.527 | 18.886 | 0.958 | 0.000 |

| Time | Coal (EJ/yr) | Oil (EJ/yr) | Gas (EJ/yr) | Nuclear (EJ/yr) | Hydro (EJ/yr) | Wind (EJ/yr) | Solar (EJ/yr) | Geothermal (EJ/yr) | Wastes (EJ/yr) | Wave/ Tidal (EJ/yr) | Modern biofuels (EJ/yr) | Woodfuel/ Crop residues (EJ/yr) | Food (EJ/yr) | Fodder (EJ/yr) | Trad. water/wind (EJ/yr) |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1987 | 81.463 | 131.805 | 72.229 | 16.503 | 8.454 | 0.003 | 0.000 | 0.800 | 0.408 | 0.004 | 0.309 | 38.933 | 19.236 | 0.947 | 0.000 |
| 1988 | 83.433 | 136.676 | 75.812 | 17.877 | 8.778 | 0.005 | 0.000 | 0.816 | 0.432 | 0.004 | 0.315 | 39.339 | 19.591 | 0.945 | 0.000 |
| 1989 | 84.635 | 139.062 | 78.367 | 18.457 | 8.724 | 0.037 | 0.006 | 0.986 | 0.876 | 0.004 | 0.321 | 39.745 | 19.944 | 0.930 | 0.000 |
| 1990 | 86.008 | 141.008 | 80.233 | 19.240 | 9.083 | 0.051 | 0.010 | 1.075 | 1.017 | 0.004 | 0.313 | 40.151 | 20.291 | 0.911 | 0.000 |
| 1991 | 78.211 | 140.804 | 81.243 | 21.776 | 9.246 | 0.059 | 0.012 | 1.112 | 1.113 | 0.004 | 0.343 | 40.597 | 20.629 | 0.868 | 0.000 |
| 1992 | 77.788 | 141.256 | 81.467 | 21.988 | 9.268 | 0.066 | 0.011 | 1.146 | 1.228 | 0.004 | 0.331 | 41.043 | 20.961 | 0.835 | 0.000 |
| 1993 | 75.864 | 142.421 | 82.921 | 22.709 | 9.799 | 0.080 | 0.013 | 1.175 | 1.265 | 0.004 | 0.335 | 41.490 | 21.286 | 0.804 | 0.000 |
| 1994 | 77.633 | 140.224 | 83.719 | 23.184 | 9.896 | 0.105 | 0.014 | 1.171 | 1.333 | 0.004 | 0.371 | 41.936 | 21.604 | 0.771 | 0.000 |
| 1995 | 79.776 | 143.462 | 84.853 | 24.110 | 10.390 | 0.114 | 0.015 | 1.148 | 1.424 | 0.004 | 0.381 | 42.382 | 21.916 | 0.751 | 0.000 |
| 1996 | 81.357 | 146.908 | 87.371 | 24.999 | 10.545 | 0.134 | 0.017 | 1.221 | 1.449 | 0.004 | 0.379 | 42.828 | 22.223 | 0.730 | 0.000 |
| 1997 | 85.999 | 151.811 | 87.390 | 24.778 | 10.782 | 0.175 | 0.018 | 1.264 | 1.556 | 0.004 | 0.435 | 43.274 | 22.524 | 0.709 | 0.000 |
| 1998 | 86.658 | 154.685 | 89.070 | 25.266 | 10.803 | 0.232 | 0.019 | 1.341 | 1.614 | 0.004 | 0.415 | 43.721 | 22.821 | 0.680 | 0.000 |
| 1999 | 85.894 | 152.612 | 91.173 | 26.107 | 10.978 | 0.306 | 0.021 | 1.436 | 1.693 | 0.004 | 0.398 | 44.167 | 23.117 | 0.642 | 0.000 |
| 2000 | 87.231 | 159.317 | 94.579 | 26.726 | 11.107 | 0.452 | 0.026 | 1.547 | 1.773 | 0.004 | 1.154 | 44.613 | 23.412 | 0.622 | 0.000 |
| 2001 | 91.515 | 159.044 | 96.584 | 27.457 | 10.875 | 0.552 | 0.033 | 1.535 | 1.868 | 0.004 | 1.184 | 44.661 | 23.709 | 0.589 | 0.000 |
| 2002 | 94.241 | 157.748 | 98.388 | 27.770 | 11.018 | 0.760 | 0.041 | 1.559 | 2.030 | 0.004 | 1.250 | 44.710 | 24.007 | 0.553 | 0.000 |
| 2003 | 100.082 | 163.184 | 101.626 | 27.466 | 11.056 | 0.926 | 0.052 | 1.610 | 2.137 | 0.004 | 1.356 | 44.758 | 24.307 | 0.527 | 0.000 |
| 2004 | 105.911 | 171.175 | 104.302 | 28.570 | 11.782 | 1.212 | 0.066 | 1.675 | 2.280 | 0.004 | 1.410 | 44.807 | 24.609 | 0.503 | 0.000 |
| 2005 | 117.381 | 174.723 | 107.318 | 28.639 | 12.305 | 1.498 | 0.091 | 1.698 | 2.445 | 0.004 | 1.498 | 44.855 | 24.914 | 0.476 | 0.000 |
| 2006 | 123.856 | 174.983 | 111.258 | 29.019 | 12.753 | 1.898 | 0.124 | 1.740 | 2.599 | 0.004 | 1.647 | 44.903 | 25.222 | 0.439 | 0.000 |
| 2007 | 129.046 | 174.443 | 113.407 | 28.454 | 12.927 | 2.456 | 0.166 | 1.819 | 2.772 | 0.004 | 1.873 | 44.952 | 25.533 | 0.397 | 0.000 |
| 2008 | 133.355 | 177.017 | 117.342 | 28.339 | 13.469 | 3.172 | 0.273 | 1.902 | 2.860 | 0.004 | 2.169 | 45.000 | 25.847 | 0.356 | 0.000 |
| 2009 | 136.117 | 175.391 | 114.711 | 27.931 | 13.697 | 3.975 | 0.463 | 1.966 | 3.081 | 0.004 | 2.223 | 45.048 | 26.163 | 0.306 | 0.000 |
| 2010 | 144.315 | 183.221 | 122.908 | 28.688 | 14.493 | 4.919 | 0.747 | 1.989 | 3.519 | 0.004 | 2.405 | 45.097 | 26.481 | 0.269 | 0.000 |
| 2011 | 152.336 | 184.703 | 127.047 | 27.466 | 14.777 | 6.428 | 1.452 | 2.026 | 3.886 | 0.004 | 2.319 | 45.145 | 26.801 | 0.222 | 0.000 |
| 2012 | 148.757 | 189.760 | 130.399 | 25.580 | 15.442 | 7.487 | 2.301 | 2.201 | 4.150 | 0.004 | 2.188 | 45.194 | 27.123 | 0.169 | 0.000 |
| 2013 | 150.034 | 190.615 | 132.289 | 25.705 | 15.676 | 9.152 | 3.047 | 2.321 | 4.459 | 0.004 | 2.219 | 45.242 | 27.444 | 0.116 | 0.000 |
| 2014 | 149.432 | 195.082 | 135.141 | 26.371 | 15.606 | 10.084 | 4.016 | 2.477 | 4.820 | 0.004 | 2.190 | 45.290 | 27.766 | 0.062 | 0.000 |

# APPENDIX C: COST SHARE THEOREM[1]

Entrepreneurial decisions, aiming at producing a certain quantity of output $Y$ within the technology that exists at a given time $t$, determine the absolute magnitude of the total capital stock, its degree of capacity utilization, and its degree of automation. The machines of the capital stock are designed and built for specific energy inputs and require a certain amount of labor for handling, supervision, and maintenance. The quantities of labor and energy that are combined with the capital stock of a fixed degree of automation determine the degree of capacity utilization. The degree of automation at time $t$ is represented by the ratio of the actual capital stock $K$ to the capital stock $K_m(Y)$ that would be required in order to produce the actual output $Y$ with the actual technology in the state of maximum automation. This state is characterized by a combination of capital and energy such that adding one more unskilled worker adds virtually nothing to gross economic output so that the output elasticity of routine labor would be vanishingly small. In some manufacturing sectors of industrialized countries this point actually does not seem to be far away.

It is obvious from an engineering point of view and by definition that both the degree of capacity utilization and the degree of automation (i) are functions of capital, labor and energy and (ii) cannot exceed the number 1. (In fact, even these days, after 40 years during which the density of transistors on a microchip has doubled every 18 months, the achievable state of automation of an economy is well below 1.) In other words, a production system cannot operate above design capacity, and the maximum degree of automation cannot be exceeded. These are the two fundamental technological constraints on the combinations of capital, labor and energy in modern economies. They drastically change the conditions for economic equilibrium that result from the behavioral assumptions of standard economics. One such assumption is profit maximization, according to which the actions of all economic agents are supposed to move the economy into a point of factor space where the difference between output and total factor cost is maximum. Alternatively one may follow Samuelson & Solow (1956) and assume that: "… society maximizes the (undiscounted) integral of all future utilities of consumption subject to the fact that the sum of current consumption and of current capital formation is limited by what the current capital stock can produce." The optimization calculus according to these two extremum principles is presented in Kümmel et al. (2010). The following is the summary of its results. For the sake of notational convenience we write the production factors $K$, $L$, and $E$ as $X_1$, $X_2$, $X_3$ and the output elasticities $\varepsilon_i$, with $i = 1, 2, 3$ defined as

$$\varepsilon_i \equiv \frac{X_i}{Y}\frac{\partial Y}{\partial X_i}. \qquad \text{(C.1)}$$

Let us consider an economic system that produces its output $y$ with three factors of production $X_1$, $X_2$, $X_3$, whose combinations are subject to technological constraints, labeled by the indices A and B and expressed by the equations $f_A(X_1, X_2, X_3, t) = 0$, $f_B(X_1, X_2, X_3, t) = 0$ with the help of slack variables. They concern the degree of capacity utilization and the degree of

---

[1] The following demonstration of the inaccuracy of the so-called cost share theorem is entirely reproduced from Lindenberger & Kümmel (2011).

automation. Their explicit forms are given in Kümmel et al. (2010). Then, profit maximization under constant returns to scale results in the three equilibrium conditions

$$\varepsilon_i = \frac{X_i(p_i + s_i)}{\sum_{i=1}^{3} X_i(p_i + s_i)}, \qquad i = 1, 2, 3, \tag{C.2}$$

which relate the output elasticities $\varepsilon_i$ of factors $X_i$ to the market prices $p_i$ per factor unit and the factor shadow prices

$$s_i \equiv -\mu_A \frac{\partial f_A}{\partial X_i} - \mu_B \frac{\partial f_B}{\partial X_i}. \tag{C.3}$$

Here, $\mu_A$ and $\mu_B$ are the Lagrange multipliers of the two technological constraint equations in the optimization calculus. Thus, the output elasticities in (C.2) are equal to *shadowed* cost shares. Intertemporal optimization of utility $U$ as a function of consumption $C$ yields that the shadow price of capital contains an additional term proportional to the time derivative of $dU/dC$. This term is small for weakly decreasing $U(C)$.

If there were no technological constraints, the Lagrange multipliers would be zero, and the equilibrium conditions would read

$$\frac{\partial Y}{\partial K} = p_k, \qquad \frac{\partial Y}{\partial L} = p_l, \qquad \frac{\partial Y}{\partial E} = p_e. \tag{C.4}$$

The shadow prices $s_i$ would vanish, and one would have the usual factor cost shares on the r.h.s of equation (C.2). That’s why standard economics assumes that in economic equilibrium output elasticities equal factor cost shares. As shown in Kümmel et al. (2010), this would also justify the duality of production factors and factor prices, which is often used in orthodox growth analyses. The essential information on production would be contained in the price function as the Legendre transform of the production function. In the presence of technological constraints and non-zero shadow prices, however, the Lagrange multipliers are finite and functions of the output elasticities $\varepsilon_i$, so that the cost share theorem and duality are not valid. For an understanding of the economy, prices are not enough

## APPENDIX D: FACTORS OF PRODUCTION AND LINEX PRODUCTION FUNCTION

### *Defining factors of production on a common ground*

In order to deal with both the inaccurate negligence of energy in production function and the intrinsic critics of the very notion of production function, Kümmel has searched for a definition of production factors aggregated in what shall be regarded as an accurate production function respecting technological constraints. In this model, the heterogeneous productive factors are aggregated according to the relevant characteristics that they all have in common: *work performance* and *information processing*. Indeed, these two features can characterize every purposeful activity, as highlighted in Kümmel et al. (2002): “energy conversion in the human body and in the machines and other energy conversion devices of the capital stock provides the work that (i) moves masses and erects structures, (ii) arranges the atoms of the raw materials in the order required for the finished products, and (iii) drives the electrons in the electric devices of the goods-producing and service industries. Information in the form of energy signals and their modulation, processed in the brain and nervous system of humans and in the transistors and other switching devices of the capital stock, controls the flow of work during the production of goods and services”. Thus in this model, capital $K$, labor $L$ and energy $E$ are the productive factors that generate the output $Y$ through their work performance and information processing. More precisely:

- The output $Y$ is a homogenous industrial/service output that should be regarded as the technological aggregation of the physical work performed and the number of information units processed in its generation. The averaged product of these two quantities is proportional to the real monetary value of $Y$ present in the national accounts in constant currency.
- Capital stock $K$ consists in all energy-conversion devices and the installations and buildings necessary for their operation and maintenance. $K$ is a technological aggregation of the maximum amount of work performed and information processed per unit time of all its constituents[1]. Here again the averaged product of capital potential of work performance and information processing is proportional to the real monetary value of $K$ shown in the national accounts in constant currency.
- Labor $L$ is defined as the aggregation of all actually worked man-hours per year necessary to manipulate and supervise the capital stock $K$ and the flow of energy $E$.

---

[1] The precise definition of the unit of capital $Y$ and output $Y$ is given in Kümmel (1982) and summarized here following Kümmel et al. (2002). “Capital is measured in ATON (for AuTomatiOn), where 1 ATON = 1kW× $\kappa$ kilobits/s. The average equivalence factor $\kappa$ is given by $\kappa = (1/N) \sum_{i=1}^{N} S_i T_i$, where the definitions of $N$, $S_i$ and $T_i$ imply the measurement prescription of the ATON: $K$ is partitioned in $N \gg 1$ pieces $K_i$, which all have the same monetary value, say $\nu$ EUROs. Then, $S_i =$ number of kilowatts performed, and $T_i =$ number of kilobits/s processed by the fully employed $i$th capital good $K_i$. As a consequence of these definitions, the ATON value of $K$, $A_k$, is proportional to the monetary value of $K$, $M_k$, shown in the national accounts in constant currency, as long as $\kappa$ stays constant: $A_k \equiv N\ ATONs = \sum_{i=1}^{N} S_i T_i$ kW × kB/s. $M_k \equiv N\nu$ EUROs, thus $A_k = M_k/\nu$. (ATON/EURO). Changes of $\kappa$ occur when the monetary valuation of the capabilities of work performance and information processing changes. The capital services of performed work and processed information flow from the capital stock to the same extent as energy and labor activate and control that stock. An equivalence factor $\zeta$, similar to $\kappa$, appears in the technical definition of output $Y$ in terms of the physical work performed and the number of information units processed in its generation.”

- Energy $E$ is the aggregation of all the energy quantities (e.g. Joules) consumed during one year in order for labor $L$ to activate the capital stock $K$.

If we introduce time *t* as the growth interval within which technology is necessary given, we ensure that the factor of proportionality between the technological and the monetary $Y$- and $K$-quantities is also constant within $t$. Furthermore, we have to assume that capital $K$, labor $L$ and energy $E$ are independent variables in the sense that within technological limits producers can vary them independently according to their prior decisions on the capital stock's quantity, degree of automation (i.e. quality) and degree of capacity utilization. In other words, it means that "variations of labor and energy at constant quantity and quality of capital are associated with variations in the degree of utilization, whereas changes in automation change the relative magnitude of the labor and energy inputs that are required to handle and activate capital at a given degree of capital utilization" (Kümmel et al., 2002). We will also assume that within a growth interval, the output $Y(K, L, E)$ is a unique, analytic, twice differentiable function of the production factors.

Given the previous definitions and assumptions, it is easy to explain why land and raw materials are not considered as production factors in Kümmel's model. Indeed, even if in preindustrial agrarian societies land was the source of economic and political power and thus was considered as the most important production factor, production actually occurred thanks to plant photosynthesis and the work performed and information processed by the people and animals tilling the soil. Therefore, land, or rather its three-dimensional extension *space* (i.e. biosphere), is a production site but not an active factor as long as its capacity of absorption of polluting emissions is not binding (Kümmel et al., 2010). In the same way, raw materials neither perform work nor process information. Raw materials remain passive during the production process where "their atoms and electrons are rearranged by capital, labor and energy into the configurations of flows required [to generate] a product or service" (Kümmel et al., 2002). Hence, raw materials do not contribute actively to the generation of value added and can consequently be ignored as long as their finiteness nature does not constraint growth.[1]

Finally, the last factor that is incorporated in the production function of Kümmel's model is *creativity*. This factor represents the specific human intellectual contribution to the economic process through all the ideas, inventions and knowledge in the possession of humankind. Contrary to the other three explicit production factors, creativity is more discreet as it is expressed through the time dependence of the production function defined by Kümmel as accurate and that we are going to present now.

*Linex production function*

Usually, a modeler arbitrarily chooses the form of the function that represents output production in an aggregated macroeconomic model. Most commonly used functions are of constant elasticity of substitutions (CES) types, with the Cobb-Douglas function as a particular case quite frequently employed in models. These functions have characteristics that enable quite

---

[1] To the people arguing that material scarcity could indeed induces constraints on growth, one may respond that *quantitative scarcity* is not a real issue since all it takes to extract material from increasingly low grade deposits is "only" increasing energy, thus what really constraint growth is *qualitative scarcity*, as depicted in Chapter 5.

easy mathematical manipulation, but in contrast they necessarily present determinate features that impose strong assumptions: in CES functions, both output elasticities[1] and the elasticity of input substitution[2] are constant and the later can take a value between 0 and positive infinite; in the Cobb-Douglas particular case, output elasticities are also constant and the elasticity of input substitution is necessarily equal to one, meaning that production factors are substitutable to the extent that neither of them is essential. Here, we will choose a different approach that starts from the usual differential equations based on the causal relations between production factors and output. Then we will define technological boundary conditions that will lead to the so-called Linex function developed by Kümmel (1982), which depends *linearly* on energy and *exponentially* on the ratios of labor to capital and energy to capital. The Linex production function is also named capital-labor-energy-creativity (KLEC) model.

An infinitesimal increase $dY$ of output is related to the infinitesimal increments $dK$, $dL$, and $dE$ of the production factors by

$$dY = \frac{\partial Y}{\partial K} dK + \frac{\partial Y}{\partial L} dL + \frac{\partial Y}{\partial E} dE + \frac{\partial Y}{\partial t} dt. \tag{D.1}$$

Since $\partial Y$ is a total differential, expression (D.1) can be rewritten as

$$\frac{dY}{Y} = \frac{K}{Y}\frac{\partial Y}{\partial K}\frac{dK}{K} + \frac{L}{Y}\frac{\partial Y}{\partial L}\frac{dL}{L} + \frac{E}{Y}\frac{\partial Y}{\partial E}\frac{dE}{E} + \frac{t-t_0}{Y}\frac{\partial Y}{\partial t}\frac{dt}{t-t_0}. \tag{D.2}$$

We can define the following output elasticities of capital ($\alpha$), labor ($\beta$) and energy ($\gamma$) which represent the respective weights with which the growth rates of the production factors contribute to the growth of the output production. Similarly, creativity in the previous section is represented by $\delta$.

$$\alpha \equiv \frac{K}{Y}\frac{\partial Y}{\partial K}, \beta \equiv \frac{L}{Y}\frac{\partial Y}{\partial L}, \gamma \equiv \frac{E}{Y}\frac{\partial Y}{\partial E}, \delta \equiv \frac{t-t_0}{y}\frac{\partial Y}{\partial K}. \tag{D.3}$$

Hence, we find the well-known growth equation

$$\frac{dY}{Y} = \alpha\frac{dK}{K} + \beta\frac{dL}{L} + \gamma\frac{dE}{E} + \delta\frac{\partial Y}{\partial t}\frac{dt}{t-t_0}. \tag{D.4}$$

Within the growth interval $t$, the contributions of all production factors must satisfy constant returns to scale expressed through

$$\alpha + \beta + \gamma = 1, \tag{D.5}$$

---

[1] The output elasticity of a production factor $X_i$ is the percentage change in output $Y$ per one percent change in input factor $X_i$, all other production factors remaining constant.

[2] The elasticity of input substitution $\sigma_{ij}$ between the two production factors $X_i$ and $X_j$ gives the ratio of the relative change of factor quotients to the relative change of the quotients of the marginal productivities, if only factors $X_i$ and $X_j$ vary and all other factors stay constant (Lindenberger & Kümmel 2011).

this ensures that our production function is linearly homogenous (i.e. homogenous of degree one). In addition, since technical-economic rationality implies that the increase in an input factor cannot results in a decrease of the output, output elasticities are necessarily non-negative as described by

$$\alpha \geq 0\,, \qquad \beta \geq 0\,, \qquad \gamma = 1 - \alpha + \beta \geq 0. \tag{D.6}$$

In order to comply with twice differentiability and path-independent integrability, the second order derivatives of $Y[K, L, E; t]$ must be equal:

$$\frac{\partial^2 Y}{\partial K \partial L} = \frac{\partial^2 Y}{\partial L \partial K}, \frac{\partial^2 Y}{\partial K \partial E} = \frac{\partial^2 Y}{\partial E \partial K}, \frac{\partial^2 Y}{\partial L \partial E} = \frac{\partial^2 Y}{\partial E \partial L}. \tag{D.7}$$

As shown in Lindenberger & Kümmel (2011), the combination of the symmetry restrictions (D.7) with (D.1) yields the following set of partial differential equations:

$$\begin{gathered} K\frac{\partial \alpha}{\partial K} + L\frac{\partial \alpha}{\partial L} + E\frac{\partial \alpha}{\partial E} = 0, \\ K\frac{\partial \beta}{\partial K} + L\frac{\partial \beta}{\partial L} + E\frac{\partial \beta}{\partial E} = 0, \\ L\frac{\partial \alpha}{\partial L} = K\frac{\partial \beta}{\partial K}. \end{gathered} \tag{D.8}$$

According to Kümmel (1982), the most general solutions of the first and second differential equation of (D.8) are given by

$$\alpha = \mathrm{A}\left(\frac{L}{K}, \frac{E}{K}\right), \qquad \beta = \mathrm{B}\left(\frac{L}{K}, \frac{E}{K}\right), \qquad and\ \gamma = 1 - \alpha - \beta. \tag{D.9}$$

Where A and B are any continuous and differentiable functions of their arguments $L/K$ and $E/K$. They are necessarily associated by (D.10) because of the third equation in (D.8). Hence, with $J(L/E)$ as any differentiable function of $L/E$, (D.10) is as follows

$$\beta = \int^{K} \frac{L}{K'} \frac{\partial A}{\partial L} dK' + J\left(\frac{L}{E}\right). \tag{D.10}$$

Following Lindenberger & Kümmel (2011), the general form of the twice differentiable, linearly homogenous production function with output elasticities (D.9) is thus described by

$$Y = E\mathcal{F}\left(\frac{L}{K}, \frac{E}{K}\right). \tag{D.11}$$

According to the theory of partial differential equations, this production function could be uniquely determined if one could define $\beta$ on a boundary surface of $K$, $L$, $E$-space, and $\alpha$ on a boundary curve in that space (Kümmel, 1982). Unfortunately, since we are lacking such information on the technical-economical state of the system, we are going to use less mathematical stringent conditions. For this, we have to postulates that $\alpha$ and $\beta$, both functions of $K$, $L$, $E$ present asymptotic technological boundary conditions. First, capital cannot be activated without the appropriate amount of energy and labor. Thus, the weight $\alpha$ of capital contribution to the growth of output $y$ would decrease with increasing $K$, to the extent that if $L/K$ and $E/K$ go to zero, $\alpha$ should vanish. Moreover, as $L$ and $E$ are of the same physical nature, they should add in any function for $\alpha$. Second, with sufficient capital and energy, a state of maximum automation could be reached such that the addition of another unit of routine labor would not contribute to output any more. More precisely, if $K_m(Y)$ is the fully employed capital stock in the state of maximum automation, and $E_m \equiv cK_m(Y)$ is its energy demand with $c$ as the *energy efficiency* of the capital stock at the given state of technology within the growth interval, then the output elasticity of labor, $\beta$, should vanish if $K$ approaches $K_m(Y)$ and e approaches $E_m \equiv cK_m(Y)$. Hence, the simplest output elasticities that satisfy the differential equations (D.8), the constant returns to scale ensured by (D.5) and the previously described asymptotic technological boundary conditions are given

$$\alpha = a\left(\frac{L+E}{K}\right), \qquad \beta = a\left(\frac{cL}{E} - \frac{L}{K}\right), \qquad \gamma = 1 - a\frac{E}{K} - ac\frac{L}{E}. \tag{D.12}$$

Here, the parameter $a$ represents the effectiveness with which energy activates and labor handles the capital stock (i.e. $a$ is the *capital efficiency*). As can be seen in (D.10), the negative term in $\beta$ is a direct consequence of the choice of $\alpha$, whereas the positive term is due to the special choice of the function $J(L/E)$ so that the asymptotic condition for $\beta$ is respected.
If we insert (D.12) into (D.4), we obtain the equation of growth

$$\frac{dY}{Y} = a\frac{L+E}{K^2}dK + a\left(\frac{c}{E} - \frac{1}{K}\right)dL + \left(\frac{1}{E} - a\left(\frac{1}{K} + c\,\frac{L}{E^2}\right)\right)dE. \tag{D.13}$$

Finally, the exact integral of (D.13) is the so-called Linex production function (D.14), which depends linearly on energy and exponentially on quotients of capital, labor and energy:

$$Y_{LinEx} = E\exp\left(a\left(2 - \frac{L+E}{K}\right) + ac\left(\frac{L}{E} - 1\right)\right). \tag{D.14}$$

The Linex production function is said to be of VES type, for variable elasticities of substitution. Indeed, if we calculate the (Hick or direct) elasticity of substitution $\sigma_{ij}$ between factors $X_i$ and $X_j$ as defined by

$$\sigma_{ij} = -\frac{d(X_i/X_j)}{(X_i/X_j)}\left(\frac{d\left(\frac{\partial Y/X_i}{\partial Y/X_j}\right)}{\left(\frac{\partial Y/X_i}{\partial Y/X_j}\right)}\right)^{-1}. \tag{D.15}$$

We find, according to Lindenberger & Kümmel (2011), that the elasticities of substitution between capital and labor ($\sigma_{kl}$), capital and energy ($\sigma_{ke}$) and labor and energy ($\sigma_{le}$) depend on the output elasticities expressed in (D.12) and on the term $a\,L/K$:

$$\begin{aligned} \sigma_{kl} &= \frac{\alpha+\beta}{2(\beta+aL/K)}, \\ \sigma_{ke} &= \frac{(1-\beta)(1-\alpha-\beta)}{2(1-\beta)(1-\beta-aL/K)-\alpha}, \\ \sigma_{le} &= \frac{-(1-\alpha)(1-\alpha-\beta)}{\beta(1-2\alpha)+2(1-\alpha)aL/K}. \end{aligned} \tag{D.16}$$

It must be noted that if one choose the trivial solutions of constant output elasticities $\alpha = \alpha_{CD}$, $\beta = \beta_{CD}$, and $\gamma = \gamma_{CD} = 1 - \alpha_{CD} - \beta_{CD}$, and insert them into (D.5) at fixed *t*, one obtains after integration the energy-dependent Cobb-Douglas (CD) production function:

$$Y_{CD} = K^{\alpha_{CD}} L^{\beta_{CD}} E^{1-\alpha_{CD}-\beta_{CD}}. \tag{D.17}$$

However in this case, complete substitutability of the production factors is possible, meaning that the asymptotic technological boundary conditions described previously do not longer hold. These conditions are also forgotten in the more general CES production function (D.18) introduced by Arrow et al. (1961) and extended to more than two factors by Uzawa (1962):

$$Y_{CES} = (\psi K^{-\rho} \varphi L^{-\rho} (1-\psi-\varphi) E^{-\rho})^{-1/\rho}. \tag{D.18}$$

Where the parameters $\psi \in [0,1]$ and $\varphi \in [0,1]$ (with $0 \leq 1-\psi-\varphi \leq 1$) are usually called *distribution parameters* and represent the productivity of its associated production factor relatively to others. The constant $\rho \equiv 1/(\sigma-1)$ is determined by the constant elasticity of substitution $\sigma \in [0,+\infty]$, implying $\rho \geq -1$. The CES function could be easily brought into the form of (D.12) but as in the particular Cobb-Douglas case its output elasticities would not satisfy the asymptotic technological boundary conditions of (D.13). It must be noted that in the Linex production function defined in (D.14), there is no need for a *total productivity factor* (TFP) for reproducing past economic growth trend as in the Cobb-Douglas case. However, the coefficients $a$ and $c$ have to be determined by the modeler in order for the simulated GDP to fit real data. Hence, these coefficients, either constant or made time-dependent, appeared to be exogenous and the Linex production function (or KLEC model) cannot be a reliable basis for forecasting (Ayres, 2008). Furthermore, if one would like to use the Linex production function in an endogenous economic growth model, one would have to describe the complete endogenous mechanism explaining the evolution the values of coefficients $a$ and $c$.

That being said, the KLEC model has been applied to the sectors "Industries" and to the whole economy of the USA, Japan and Germany in several published articles. In the first studies (Kümmel 1982; Kümmel et al. 1985) using the Linex production function, coefficients $a$ and $c$ were constants and determined with the help of the Levenberg-Marquardt method of non-linear optimization under the constraints of non-negative elasticities of production. In more recent studies (Kümmel et al. 2002; 2010; Lindenberger & Kümmel 2011), the authors have developed smooth time-dependent increasing $a_t$ and decreasing $c_t$ functions of logistic type. These smooth time changes of the technology coefficients reflect structural changes towards less energy intensive production and more efficient combinations of capital, labor and energy. The choice of the logistic form was motivated by both mathematical logic (logistics are typical of growth in complex systems and processes of innovation diffusion) and empirical evidence. Figure D.1 reproduces the main results of Kümmel et al. (2002) where it can be seen that the theoretical GDP of the USA (1960 to 1996) and Germany (1960 to 1989) obtained with the Linex function fits very well the historical GDP of those same nations.

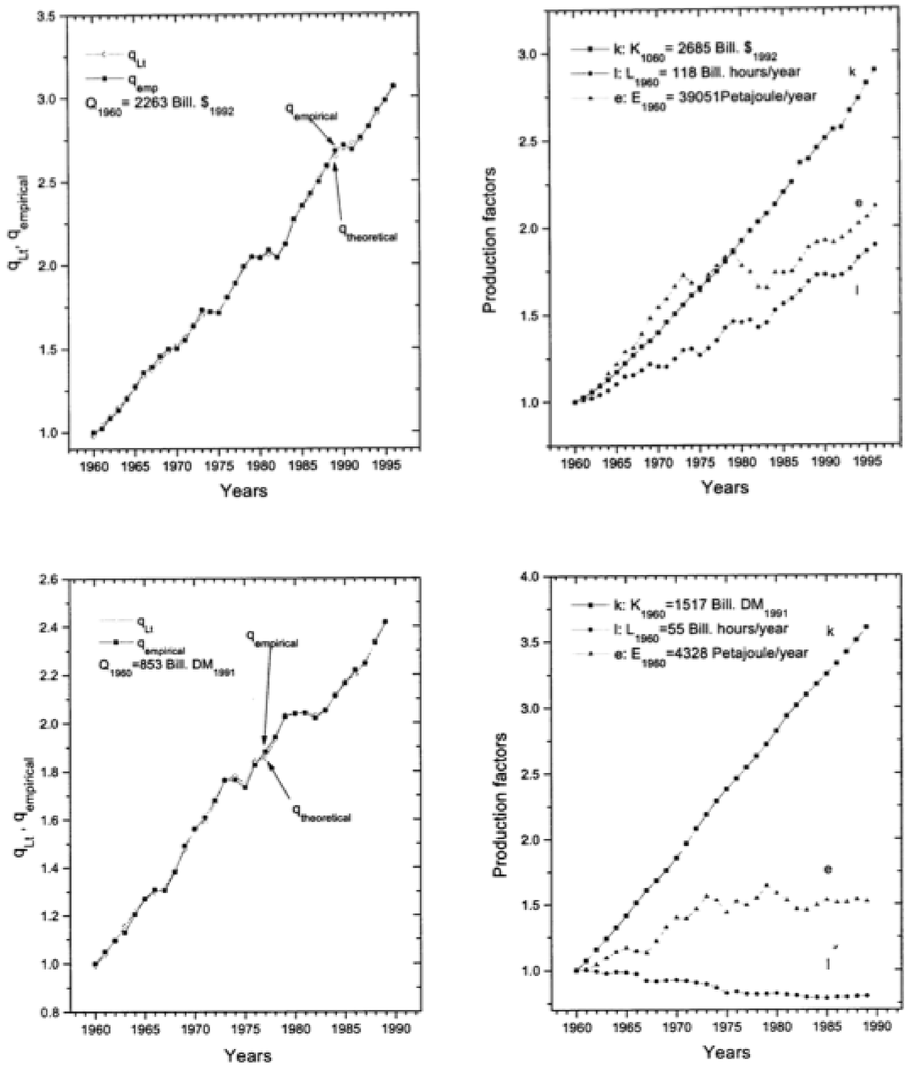


**Figure D.1 Historical and estimated GDP of the USA and Germany, 1960–1996.**
**Theoretical (diamonds) and empirical (squares) growth of annual gross domestic product $q = Q / Q_{1960}$ (left), and empirical time series of the normalized factors capital $k = K / K_{1960}$, labor $l = L / L_{1960}$ and energy $e = E / E_{1960}$ (right) in “USA, Total Economy” (top) and “Germany, Total Economy” (bottom). Source: Kümmel et al. (2002).**

Additionally, it is possible to derive from these econometric analyses the estimated average[1] output elasticities, i.e. the productive power of capital $\bar{\alpha}_{LinEx}$, labor $\bar{\beta}_{LinEx}$, energy $\bar{\gamma}_{LinEx}$ and creativity $\bar{\delta}_{LinEx}$. These average output elasticities and the statistical quality measures coefficient of determination ($R^2$) and "Durbin-Watson coefficient" ($d_w$), which best values are respectively 1 and 2, for the USA (1960 to 1996) and Germany (1960 to 1989) are reproduced in Table D.1.

**Table D.1 Calibration results of Linex production function for USA and Germany GDP.**
**Average output elasticities, coefficient of determination and Durbin-Watson coefficient for GDP calibration using the Linex production function for the USA (1960–1996) and Germany (1960–1989). Source: Kümmel et al. (2002).**

| | $\bar{\alpha}_{LinEx}$ | $\bar{\beta}_{LinEx}$ | $\bar{\gamma}_{LinEx}$ | $\bar{\delta}_{LinEx}$ | $R^2$ | $d_w$ |
|---|---|---|---|---|---|---|
| USA (1960-1996) | 0.46(±0.07) | 0.14(±0.12) | 0.30(±0.1) | 0.10(±0.08) | 0.999 | 1.91 |
| Germany (1960-1989) | 0.36(±0.03) | 0.09(±0.02) | 0.44(±0.04) | 0.11(±0.12) | 0.999 | 1.67 |

Ayres (2008) also used the Linex production function, but instead of having energy $E$ as a production factor, he used *useful exergy services (or useful work) U*. As explained in Section 3.2.3, useful work is the arithmetic product of primary exergy input $E$ with a conversion efficiency $f$ the aggregate efficiency with which primary exergy is converted into useful exergy services. Details are provided in (Ayres 2008; Ayres & Warr 2009) to explain how exactly the variables $U$ and $f$ are calculated from real. These same authors show that when output elasticities are well chosen (i.e. they do not equate total income shares), a Cobb-Douglas production function can be a fairly good approximation of the Linex production function, itself a rather good proxy of the historical GDP. Figure D.2 presents the historical and estimated GDP of the USA from 1900 to 2000, while Figure D.3 presents the same results for Japan, both from Ayres (2008). Table D.2 and D.3 respectively reproduce the statistics test results and the values of the different coefficients of the production function once fitted to empirical data.

**Table D.2 Statistics test results of Linex and Cobb-Douglas production functions.**
**Calibration to USA and Japan GDP, 1900–2005. Source: Ayres & Warr (2009).**

| | 1900–1940 | | 1949–2005 | |
|---|---|---|---|---|
| | Cobb-Douglas | Linex | Cobb-Douglas | Linex |
| *USA* | | | | |
| Dublin-Watson | 0.59 | 1.72 | 0.03 | 0.15 |
| Dickey-Fuller | -1.816$^{*}$ | -5.427$^{***}$ | 3.540 | 2.306 |
| $R^2$ | 0.987 | 0.994 | 0.997 | 0.999 |
| *Japan* | | | | |
| Dublin-Watson | 0.55 | 0.96 | 0.11 | 1.10 |
| Dickey-Fuller | -1.317 | -3.162$^{***}$ | -1.451 | -4.355$^{***}$ |
| $R^2$ | 0.985 | 0.991 | 0.999 | 1.000 |

Critical test values for the Dickey-Fuller unit-root test $^{*}$90%–1.606, $^{**}$95%–1.950, $^{***}$99%–2.366

---

[1] Since output elasticities are changing over time in the Linex production function, one must necessarily perform an average of time series in order to have something comparable to the usual Cobb-Douglas output elasticities.

**Table D.3 Parameters values of Cobb-Douglas and Linex production functions. Calibration to USA and Japan GDP, 1900–2005. Source: Ayres & Warr (2009).**

| | Coefficients of Cobb-Douglas models | | |
|---|---|---|---|
| | Capital ($\alpha$) | Labor ($\beta$) | Useful work ($1-\alpha-\beta$) |
| *USA* | | | |
| 1900-1940 | 0.33±0.064 | 0.31±0.038 | 0.35 |
| 1949-2000 | 0.78±0.037 | -0.03±0.018 | 0.25 |
| *Japan* | | | |
| 1900-1940 | 0.37±0.094 | 0.44±0.033 | 0.19 |
| 1949-2000 | 0.51±0.038 | 0.34±0.009 | 0.15 |

| | Coefficients of logistic-type parameters for $a(t)$ and $c(t)$ in the LinEx models | | | |
|---|---|---|---|---|
| | $k$ | $p$ | $q$ | $r$ |
| *USA (1900-1940)* | | | | |
| $a(t)$ | 0.08 | 97.86 | 10.26 | |
| $c(t)$ | -4.12 | 80.85 | 63.04 | 2.60 |
| *USA (1949-2000)* | | | | |
| $a(t)$ | 0.19 | 107.60 | 11.50 | |
| $c(t)$ | -0.27 | 53.44 | 89.10 | 0.47 |
| *Japan (1900-1940)* | | | | |
| $a(t)$ | 0.13 | 74.24 | 6.38 | |
| $c(t)$ | -0.06E-06 | 80.88 | 62.80 | 1.17 |
| *Japan (1949-2000)* | | | | |
| $a(t)$ | 4.53 | 233.62 | 202.74 | |
| $c(t)$ | -0.23 | 18.02 | 84.88 | 0.69 |

Where

$$a(t) = \frac{k}{1 + \exp\left[\frac{\ln(81)}{p} * (time - 1900 - q)\right]}$$

$$c(t) = \frac{k}{1 + \exp\left[\frac{\ln(81)}{p} * (time - 1900 - q) + r\right]}$$

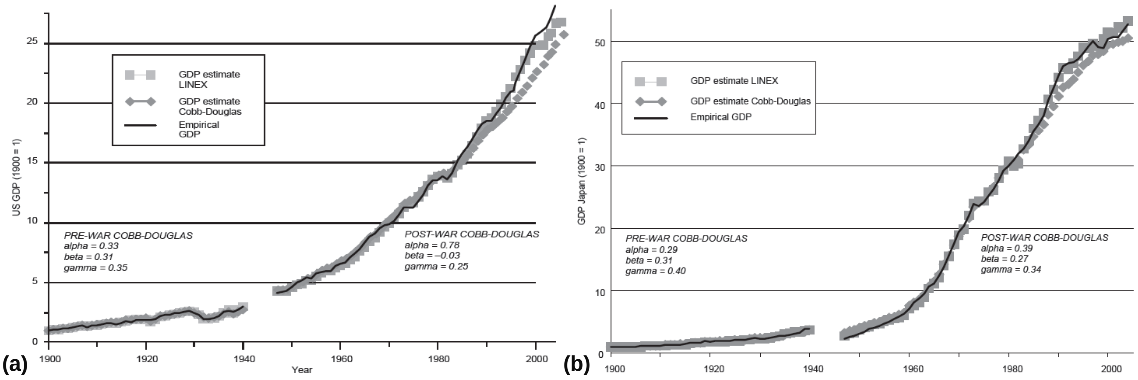
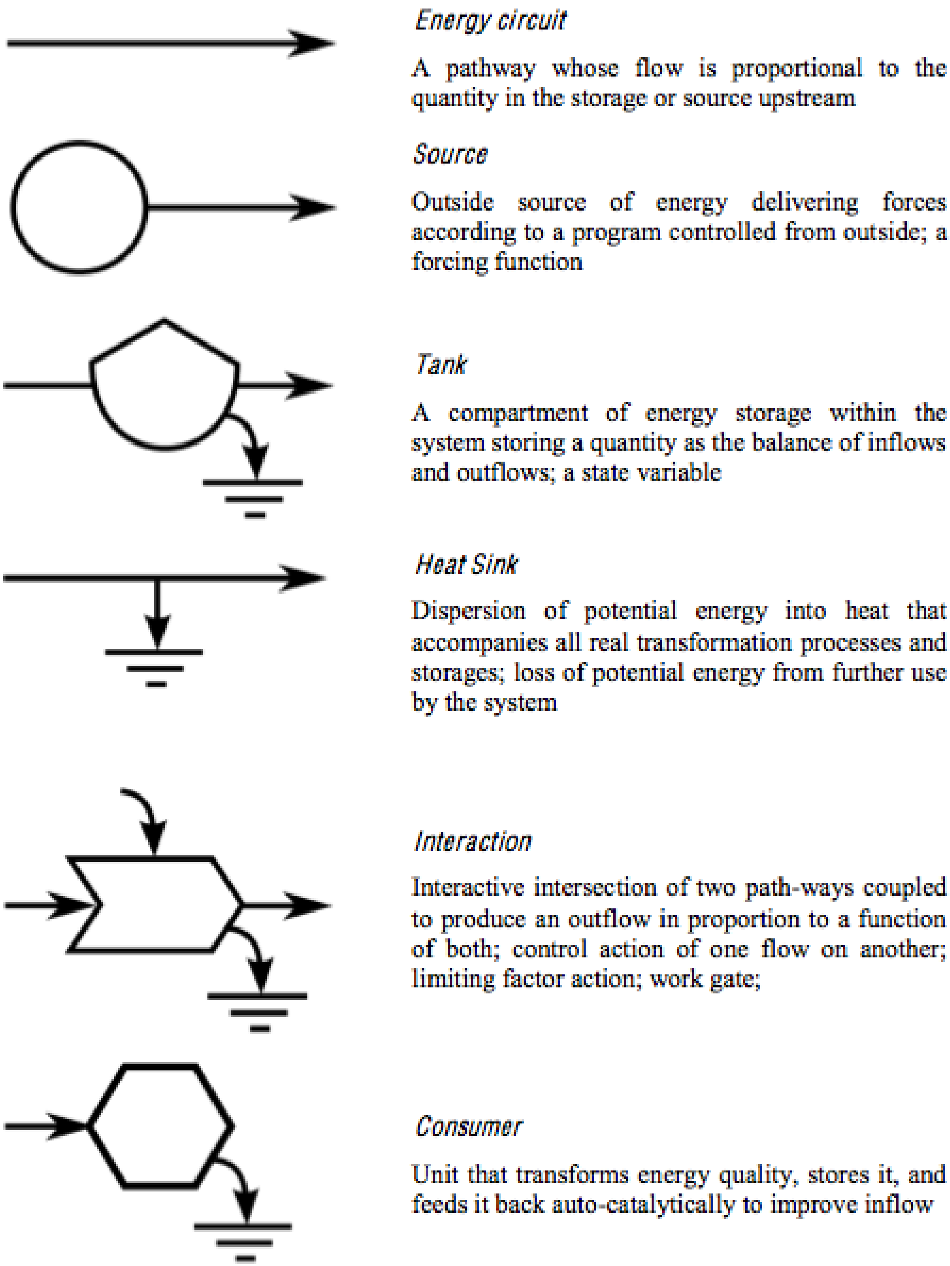


**Figure D.2 Historical and estimated GDP of the USA and Japan, 1900–2005. (a) USA and (b) Japan, both excluding 1941–1948. Source: Ayres & Warr (2009).**

What must first be noted from theses simulations is that when useful work (useful exergy services) $U$ is incorporated in the Cobb-Douglas production function, it is possible to account quite well for economic growth without any requirement for an exogenous time-dependent multiplier (the traditional total factor productivity, TFP, usually noted $A$) whereas this is impossible in the traditional capital-labor Cobb-Douglas formulation of neoclassical economics. Then, a careful reader will notice that, especially in the case of the USA, both Linex and Cobb-Douglas production functions show a growing deviation between historical and estimated GDP after 1990. Ayres (2008) attributes this gap to the recent contribution of information and computer technology (ICT) to economic growth. This argument is actually developed in a more recent work of the same authors (see Warr & Ayres 2012).

These results support the idea that useful exergy consumption is a primordial factor of production, probably even more important to economic growth than capital and labor. Indeed, these results indicate that if a Cobb-Douglas production function is to be used to represent the macroeconomic behavior of an industrialized economy, the average output elasticities representative of such an economy should be set around $\bar{\alpha}_{CB\ corrected} = 0.30$ for capital marginal productivity, $\bar{\beta}_{CB\ corrected} = 0.20$ for labor marginal productivity, $\bar{\gamma}_{CB\ corrected} = 0.5$ for energy marginal productivity. These results, compared to the usual output elasticities used in Cobb-Douglas functions: $\bar{\alpha}_{CB\ usual} = 0.30$ and $\bar{\beta}_{CB\ usual} = 0.70$, show that it is mostly the output elasticity of labor ($\beta$) that has been overestimated in mainstream economics so far. Most of what economists have called *increasing labor productivity due to technological progress* can in fact be easily attributed to *increasing quantitative substitution for labor of increasing qualitative energy*, *i.e. increasing exergy services consumption*. In other words, the results presented here support the important role for economic growth during the past two centuries of the substitution of *useful work*, mostly delivered by fossil-fuel-powered machines, for muscle work deliver by humans and animals. As summarized by Ayres (2008) himself: "Labor, in the absence of machines and sources of power, is now nearly unproductive at the margin, at least at the macro-scale. This result holds for both the US and Japan. In effect, labor is no longer scarce. One more unskilled worker, without tools and mechanical or electrical power, adds almost nothing to the economy. This has important implications for the future. It contradicts the assertions by many politicians and pundits that a declining birthrate needs to be reversed. On the contrary, the declining birthrate is more hopeful than worrisome."

# APPENDIX E: ELEMENTS OF ODUM’S ENERGY CIRCUIT LANGUAGE



*Energy circuit*

A pathway whose flow is proportional to the quantity in the storage or source upstream

*Source*

Outside source of energy delivering forces according to a program controlled from outside; a forcing function

*Tank*

A compartment of energy storage within the system storing a quantity as the balance of inflows and outflows; a state variable

*Heat Sink*

Dispersion of potential energy into heat that accompanies all real transformation processes and storages; loss of potential energy from further use by the system

*Interaction*

Interactive intersection of two path-ways coupled to produce an outflow in proportion to a function of both; control action of one flow on another; limiting factor action; work gate;

*Consumer*

Unit that transforms energy quality, stores it, and feeds it back auto-catalytically to improve inflow

Source: Hall & Day (1990).

# APPENDIX F: CORRECTING FOR QUALITY IN EROI ESTIMATION

The economic importance of energy quality and claim that the general shift to higher quality fuels affects how much energy is needed to produce GDP have early been emphasized (Odum 1971). Quality can be defined in various ways, but it can be more generally expressed as “the relative economic usefulness per heat equivalent unit of different fuels and electricity” (Cleveland et al. 2000). Energy quality is determined by a complex set of attributes unique to each fuel such as physical scarcity, capacity to do useful work (i.e. exergy content), energy density, cleanliness, ease of storage and transport, safety, flexibility of use, cost of conversion, and so on. Because heat equivalent is just one of the many attributes of a specific fuel, it ignores the context in which the fuel is used and thus cannot explain why a consumer get more utility from 1 MJ of electricity rather than 1 MJ of coal (Cleveland et al. 2000). Following this idea, many authors (Cleveland et al. 2000; Cleveland 2005; Murphy et al. 2011) have emphasized the need for energy quality correction in EROI analysis. Both the numerator and the denominator of the EROI must be quality corrected and not simply expressed in heat equivalent units. Energy quality correction can be implemented by different methods, but it always takes the following form:

$$EROI = \frac{\sum_i \lambda_i E_{out,i}}{\sum_i \lambda_i E_{in,i}} \tag{F.1}$$

Where $\lambda_i$ is a quality factor associated to fuel $i$, possibly evolving over time if the quality correction is based on relative price changes, the other possible approach being based on the relative exergy contents.

The economic approach is based on the fact that the value of a heat equivalent of fuel should be determined by its price. Indeed, in a general equilibrium context prices and marginal productivities of price-taking producers and marginal utilities of price-taking consumers are supposed to be set simultaneously. Hence, neoclassical theory supports the idea that the price per heat equivalent of fuel should equal its marginal product value and, thus, represents its economic usefulness. In other words, the willingness to pay of consumers for fuels reflects their concern about characteristics other than heat content in energy products. An example of method allowing a price-based quality correction is the Divisia Index developed by Berndt (1978). It must be noted that in such a price-based approach the full environmental and social costs associated to each fuels are not included into the market price of fuels. These externalities should be internalized in order to observe shift in energy use and in turn a change in marginal products. In this way, some doubt about the usefulness of using a price-based system for quality correction would be removed. Because other shortcomings still remain in the price-based quality correction method, some researchers support a more physical approach using exergy properties.

Based on the second law of thermodynamics, exergy is the maximum amount of physical work that can be extracted from a given flow of energy (Section 3.2.1). As work is performed, exergy is consumed until the system is in equilibrium with its surrounding (i.e. the reference state normally chosen as the atmosphere at standard temperature and pressure).

Exergy thus provide an easy way to compare energy quality of different fuels based on physical units, but it does not encompasses the great variety of characteristics describing relative quality between fuels. Hence, exergy is interesting to assess fuel quality because it avoids using economic data and the problems associated with them. But logically, the exergy approach for quality correction cannot capture as much quality properties as prices do, even considering their shortcomings described earlier. That is why in general, despite its shortcomings, the price-based approach is recommended to perform energy quality correction (Murphy et al. 2011).

# APPENDIX G: DECLINING FUNCTION OF PHYSICAL COMPONENT H

In Section 4.1.3, the physical resource component $H$ of the theoretical EROI function $F$ is assumed to decrease to an asymptotic limit as a function of production, as shown in Figure 4.5. More precisely, the physical component declines exponentially:

$$H(\rho_j) = \exp(-\varphi_j \rho_j). \tag{4.15}$$

Where $0 < \varphi_j$ represents the constant rate of quality degradation of the energy resource $j$, and $\rho_j$ is the cumulative production in case of a nonrenewable energy resource, and annual production in case of a renewable energy resource. The justification of this functional form starts by considering that in general, resources that offer the best returns (whether financial or energetic) are exploited first and attention then turns to resources offering lower returns as production continues (Dale 2010; Dale et al. 2011). The returns offered by an energy resource will depend upon the type, formation and depth of the reserve, hostility of the environment, distance from demand centers and any necessary safety or environmental measures. The costs of production often increase exponentially with increases in these factors (Cook 1976). As a consequence, the physical component of the EROI of the resource declines exponentially as a function of production. A more precise explanation is given below.

### *The case of non-renewable resources*

The use of this exponentially declining curve is justified by considering the distribution of energy resources. Some of these resources will offer large energy returns due to some factors such as their energy density (e.g. grades of crude oil or coal), their ease of accessibility (e.g. depth of oil resources, on-shore vs. offshore), their proximity to demand centres (e.g. Texan vs. Polar oil) and possible other factors. The probability density function of the EROI of one particular source should most likely displays a positive skew, i.e. the median is less than the mean, as depicted in Figure G.1 because there are more sites with lower average grade than with higher average grade (Dale, 2010; Dale et al., 2011).

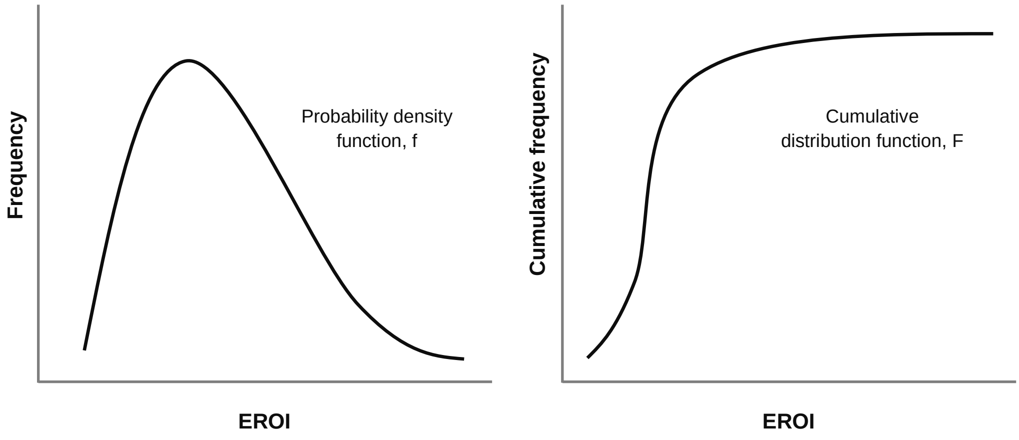


**Figure G.1 EROI probability density function and cumulative distribution function. Source: Dale, 2010.**

If we now assume that the different sites will be exploited as a function of their EROI, i.e. that those sites offering the best energy returns are exploited first, then we may now re-plot the cumulative distribution function of EROI depletion as a function of exploitation (i.e. production) by rotating the axes and ranking the sites by EROI from highest to lowest, as shown in Figure G.2 (Dale 2010; Dale et al. 2011).

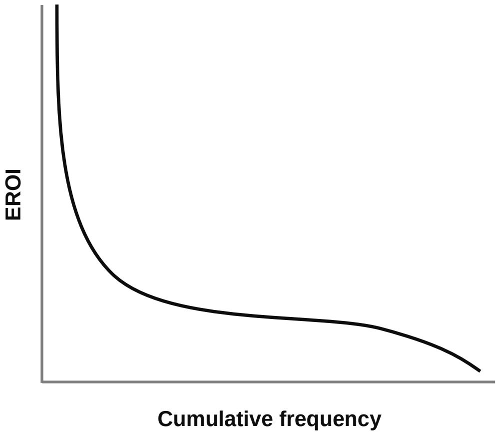


**Figure G.2 EROI as a function of the cumulative frequency of deposits/sites.**
**Source: Dale, 2010.**

### *The case of renewable resources*

The physical component $H$ of the EROI is a function of *annual* production in the case of renewable energy sources, and not *cumulative* production as in the case of nonrenewable energy resources. As a result, in case of a reduction in production, the EROI of a renewable resource may “move back up the slope” of the physical component $H$. In the meantime, technology, which is a function of cumulative production, will have increased, further pushing up energy returns. This implies that the EROI of a renewable energy source is a path dependent function of production (Dale 2010; Dale et al. 2011). Decline in the physical component of the EROI for renewable energy sources represents the likelihood of the most optimal sites being used earliest. For example, deployment of wind turbines presently occurs only in sites where the average wind speed is above some lower threshold and that are close to large demand centres to avoid the construction of large distribution networks. Over time, the availability of such optimal sites will decrease, pushing deployment into sites offering lower energy returns, which should be reflected in declining capacity factors over time (Dale 2010; Dale et al. 2011).

This assumption is backed up by the mapping of wind and solar resources in the USA. The National Renewable Energy Laboratory (NREL) has mapped wind turbines and solar panels farms all across the US. Each site is listed with its associated power density (W/m$^2$). These databases, respectively called Western Wind Dataset (NREL 2010b) and National Solar Radiation Database (NREL 2010a), were used by Dale (2010) to produce depletion curves of the US wind and solar resources. As shown in Figure G.3, this consists in ranking all sites by wind power density (W/m$^2$) and daily solar energy flow density (Wh/m$^2$/day) respectively. Figure G.3A shows that the power density of the US wind resource is indeed an exponentially declining function of land area up to 500 W/m$^2$ where the curve then decreases sharply. In the same way, Figure G.3b shows that the energy flow density of the US solar resource is also a declining exponential function of total land area (Dale 2010).

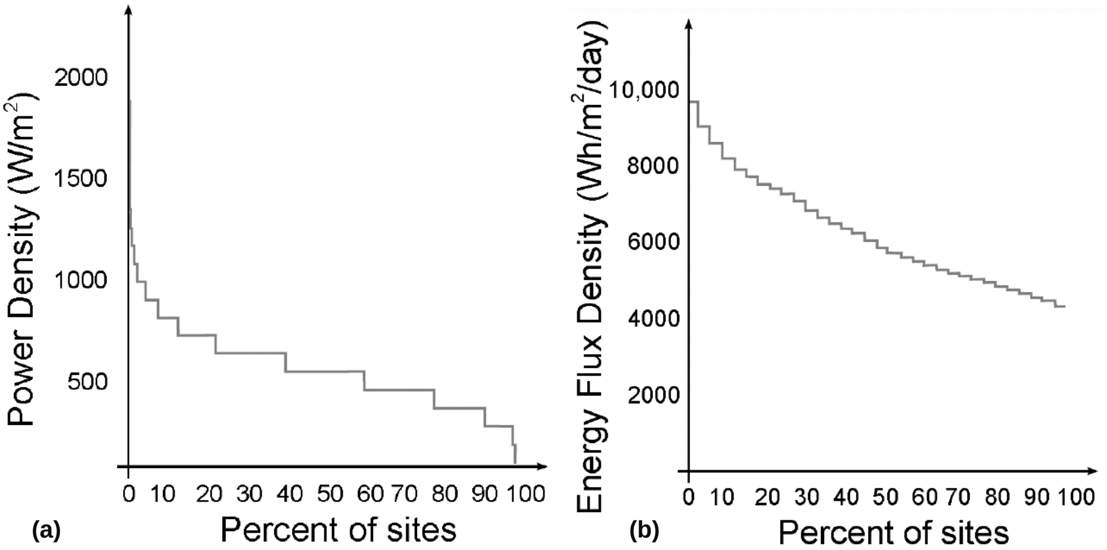


**Figure G.3 Depletion curves of wind and solar resources in the USA.**
**(a) Wind sites ranked by power density (W/m$^2$) as a percentage of total land area. (b) Solar sites Ranked by energy flow density (Wh/m$^2$/day) as a percentage of total land area. Source: Dale 2010.**

# APPENDIX H: EVOLUTION OF FOSSIL FUELS EROIs

## *Global oil and gas*

Since most of worldwide petroleum production is from national oil companies, it is extremely difficult to record reliable data for global oil and gas production. Gagnon et al. (2009) used the financial upstream database of John H. Herold Company to calculate a global EROI value for private oil and gas companies between 1992 and 2006. The results indicate an EROI for global oil and gas of approximately 23:1 in 1992, 33:1 in 1999 and 18:1 in 2006 (see Figure H.1). The maximum of 33:1 in 1999 was attained at a time of low effort in oil exploration. After 2000, oil and gas companies increased their effort in oil and gas exploration without a clear related increase in production, which remain stagnant since 2005 (for private companies). The decline of global EROI for oil and gas production will probably continue in the coming decades according to Hall et al. (2014).

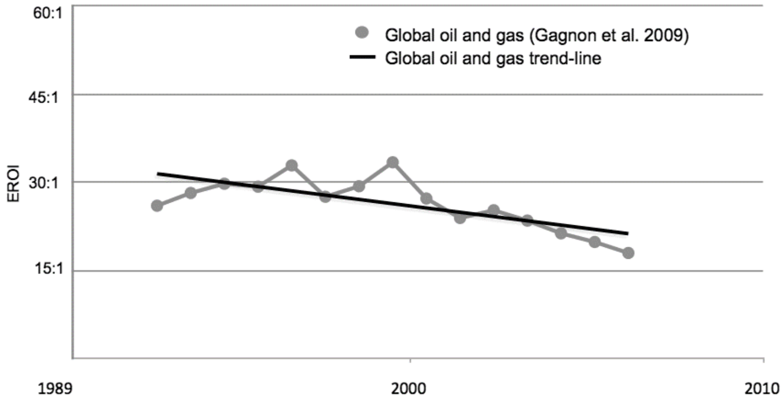


**Figure H.1 Estimates of oil and gas global EROI, 1992–2006.**
**Source: figure from Hall et al. (2014) reproducing the results of Gagnon et al. (2009).**

## *US oil and gas*

After the oil crisis of 1974 and 1975, the United States tried to improve their oil independency and thus increased their domestic drilling effort. As a result the energy investment to get oil increased but this did not lead to an increase in oil and gas production (Cleveland 2005; Guilford et al. 2011). This caused the EROI of American oil and gas production to encounter a sharp decline. Even though the drilling effort on American territory slow down at the beginning of the 90s, the EROI for oil and gas production did not increase because domestic production of conventional oil and gas continue to decrease. Guilford et al. (2011) performed the larger EROI analysis of American oil and gas production, giving estimations from 1919 to 2007. These results are presented in Figure H.2. The study of Guilford et al. (2011) was in fact the combination of two separate studies, one from Guilford & Hall, and another from Cleveland and O’Connor. Both studies show similar pattern of decreasing EROI for oil and gas production in the US with a current level estimated at 11:1. The shift from easy-to-find and exploit resources to deeper and more difficult resources explain in large part this phenomenon. Direct and indirect energy cost increased but production remained flat or decreased, so EROI has logically been declining. The intuitive idea that business cycles with higher demand for oil lead to increased prices which encourages more drilling resulting in higher oil production and then decreasing oil prices does not work in reality because more drilling does not bring more oil to the market due to geological constraints.

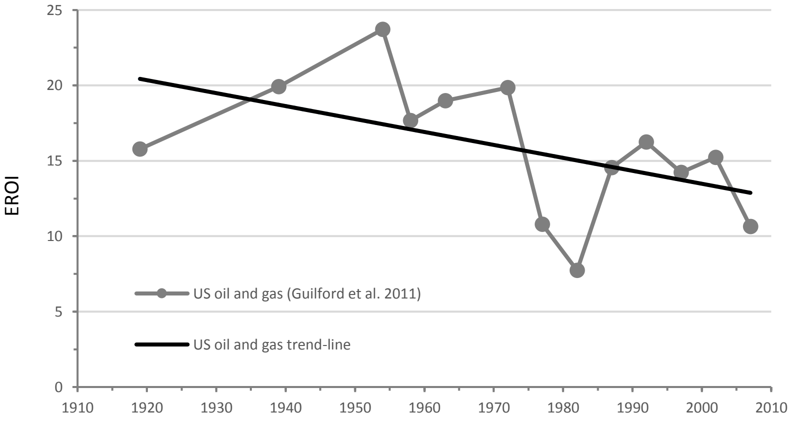


**Figure H.2 Estimates of US oil and gas EROI, 1919–2007.**
**Data source: Guilford et al. (2011).**

*Canadian oil and gas*

Freise (2011) calculated the EROI and of conventional natural gas and oil production in Western Canada (1947-2009) by a variety of methods to explore the energy dynamics of the peaking process. All these methods show a downward trend in EROI during the last decade. Natural gas EROI fell from 38:1 in 1993 to 15:1 at the peak of drilling in 2005. The drilling intensity for natural gas was so high that net energy delivered to society peaked in 2000–2002, while production did not peak until 2006. Poisson & Hall (2013) found that in Canada, the EROI of both conventional oil and gas and that of combined oil-gas-tar sands have been decreasing since the mid-1990s from roughly 20:1 to 12:1 and 14:1 to 7.5:1, or a decline of 25%. Poisson & Hall (2013) estimations of EROI for Canadian oil and gas were about half those calculated by Freise (2011), and their rate of decline is somewhat less rapid according to Lambert et al. (2012).

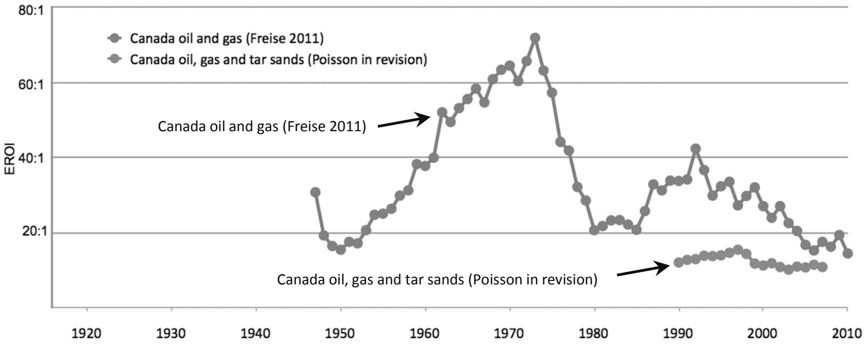


**Figure H.3 Estimates of Canadian oil and gas EROI, 1947–2009.**
**Source: figure from Hall et al. (2014) reproducing the results of Freise (2011) and Poisson & Hall (2013, shown as Poisson in revision).**

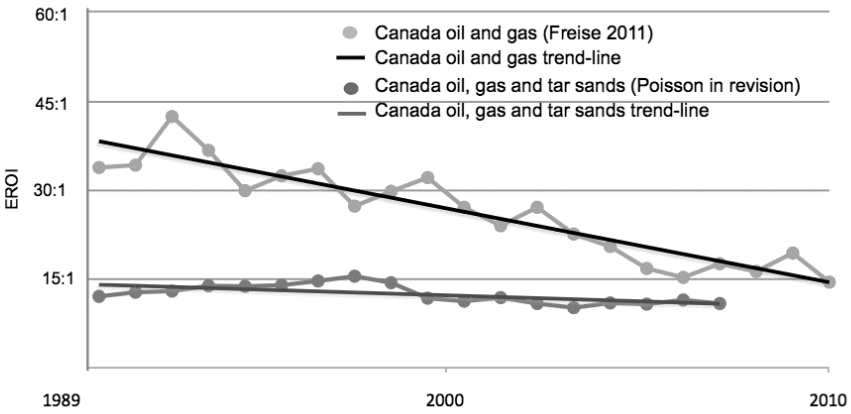


**Figure H.4 Estimates of Canadian oil and gas EROI, 1989–2009.**
**Source: figure from Hall et al. (2014) reproducing the results of Freise (2011) and Poisson & Hall (2013, shown as Poisson in revision).**

*Norwegian, Mexican and Chinese oil and gas*

The analysis performed by Grandell et al. (2011) shows that EROI for Norwegian petroleum production ranged from 44:1 in the early 1990s to a maximum of 59:1 in 1996, to about 40:1 in the latter half of the last decade. These high-energy returns on investment, higher than for the previously given data of the USA, are observed because Norwegian oil and gas fields are relatively young and of high quality. It must be noted that the declining trend in recent years is most likely due to ageing of the fields whereas varying drilling intensity might have a smaller impact on the net energy gain of the fields according to Grandell et al. (2011). These authors expect the EROI of Norwegian oil and gas production to deteriorate further as the fields become older.

In Hu et al. (2012, in fact 2013)'s paper, the EROI is derived based on existing data for production of crude oil and natural gas from the Daqing oil field, the largest oil field in China. Authors estimate that its $EROI_{stnd}$ expressed in heat equivalent was 10:1 in 2001 but has declined to 6.5:1 in 2009. Based on this trend they project that the $EROI_{stnd}$ will decline to 4.7:1 in 2015, and that the net energy from the field will be decreasing substantially (Hu et al. 2013).

Ramirez (2012, actually revised with Hall as Ramirez & Hall, 2013) analysis of Mexican oil and gas production shows a substantial decline in the past decade. This is due to the ageing of the super-giant Canterell oil field, which was the world's second largest producer of oil roughly a decade ago. According to the authors, it is not clear whether newly developed fields in this region can make up for the loss in production of Canterell field.

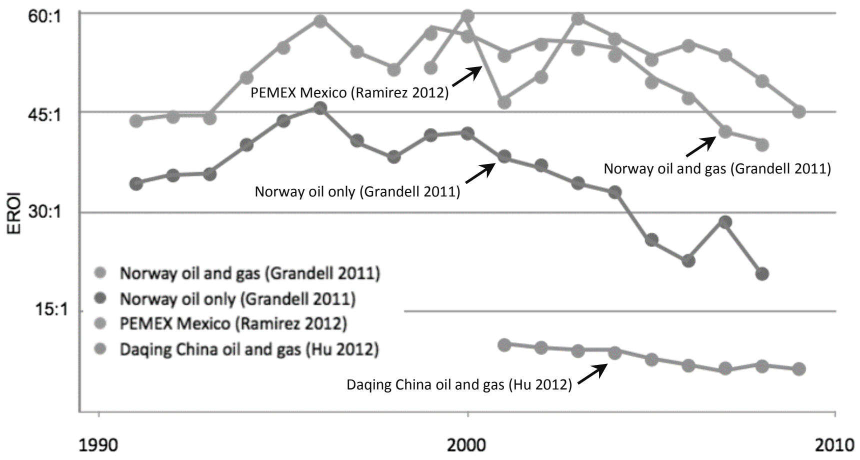


**Figure H.5 Estimates of Norwegian, Chinese and Mexican oil and gas EROI, 1989–2009.**
**Source: figure from Hall et al. (2014) reproducing the results of Grandell (2011), Hu et al. (2013, shown as Hu 2012), and Ramirez & Hall (2013 shown as Ramirez 2012).**

*American and Canadian dry gas*

Most of the time oil and gas production are associated, thus it is really difficult, if not impossible, to used the related data of such fields to try to estimate an EROI for oil and another one for gas. Only two studies have tried to assess the EROI of dry gas. The first one is from Sell et al. (2011) who examined tight natural gas deposits in the Appalachian Mountains in Pennsylvania (USA). The second is from Freise (2011) who performed an analysis of all conventional natural gas wells in western Canada. Their results are presented in Figure H.6.

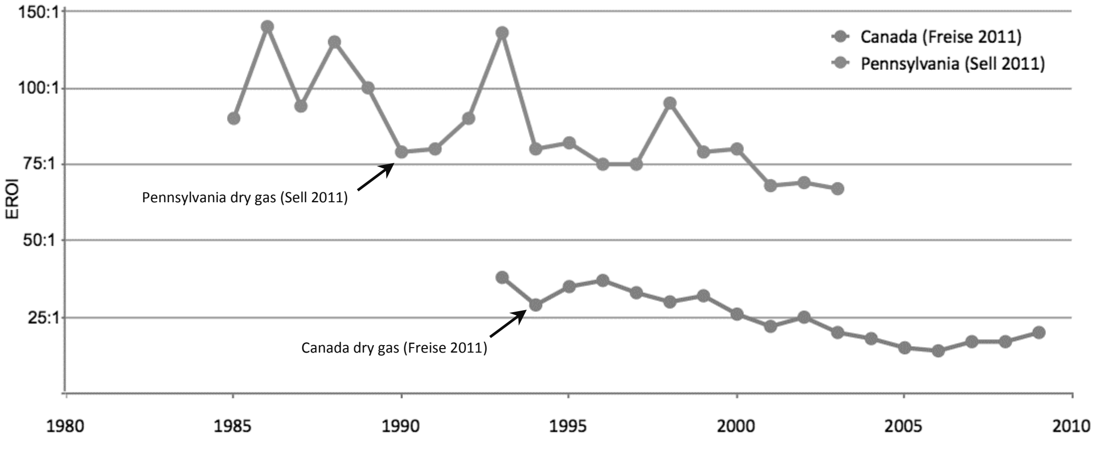


**Figure H.6 Estimates of Canadian and Pennsylvanian dry gas EROI, 1985–2009.**
**Source: figure from Hall et al. (2014) reproducing the results of Freise (2011) and Sell et al. (2011).**

*American and Chinese coal*

The only EROI estimations for coal production have been realized for the USA and China because data regarding the energy to extract coal are really scarce for other productive regions. In the USA, according to Balogh et al. (2012) analysis, the EROI of coal production declined from 80:1 in 1950 to 30:1 in 1980, and then returned to 75:1 in 2000. However, the sparse data from Cleveland (1992) are not really consistent with these results. This pattern appears to be quite strange but can if fact be simply explained. Initially, coal was mine almost exclusively in the Appalachian mountain region using a conventional method of room and pillar combination. At the beginning, the coal extracted from such a location had a high energy density (high BTUs/ton ratio) and was a mix of anthracite and high quality bituminous coal. Because the best coal was used first the EROI of coal production started to decrease slightly. As the energy investment cost increased, private firms were incentivized to look for other locations in central and northern interior states. Furthermore, the extraction method operated a shift from underground to surface mining (area, contour, auger, and mountain top mining techniques) resulting in less energy required to mine the coal. In the same time, the energy content of the coal extracted continued to decrease as the share of lower quality bituminous coal over total coal production increased. Hu et al. (2013) realized an estimation of the EROI for Chinese coal production between 1994 and 2009. Results show a stable EROI trend around 20:1, lower than in the USA case but with very less variability.

## APPENDIX I: SENSITIVITY ANALYSIS OF EROI PRICE-BASED ESTIMATES

As can been seen in equation (4.23) and (4.24), our method to estimate the global EROI of fossil fuels is logically sensitive to the uncertainty surrounding the value of its three arguments, namely:

- the prices of fossil energies presented in Figure 4.6,
- the $MROI$ supposed common to all scenarios but varying over time thanks to (4.26),
- the energy intensity $EI$ taken as the global economy average and evolving over time.

The different fossil energy prices integrate investment in energy sectors but also different kinds of rents, in particular during temporary exercise of market power. Those are not taken into account in the $MROI$ proxy. This implies that, on particular points that we cannot identify, we might have overestimated the expenditures level in a given energy sector and consequently underestimated its associated EROI. But considering that the fossil energy prices come from historical data that we consider to be robust, we think that our results are mostly subjected to the uncertainties surrounding the *MROI* and the *EI*.

### Sensitivity of price-based EROI estimates to the MROI

Regarding the estimation of the monetary-return-on-investment (MROI) in the energy sector, we propose to test two variants of the one used so far that rest on the US long-term interest rate. The total three variants are labeled A, B, and C, with the following definition:

- Variant A: the $\text{MROI}_\text{A}$ is based on the US long-term interest rate (US.LTIR) presented in Figure 4.9, to which a risk premium of 10% is added following Damodaran (2015).
- Variant B: the $\text{MROI}_\text{B}$ is based on a reconstructed AMEX Oil Index[1] based on a relation estimated between the AMEX Oil Index of Reuters (2016) for the period 1984-2012 and the NYSE Index annual returns on this same period. NYSE Index annual returns were retrieved from different references: Goetzmann et al. (2001) for the 1815-1925 period, Ibbotson & Sinquefield (1976) for the 1926-1974 period, and NYSE (2015) the 1975-2012 period (Figure I.1).
- Variant C: the $\text{MROI}_\text{C}$ is considered constant and equal to 1.1 (i.e. the energy sector gross margin is 10%). This hypothesis is the one used in previous studies such as King & Hall (2011), and King et al. (2015a).

We summarize in Table I.1 the different relations employed to estimate the MROI supposed equal (for a given year) in all the different fossil energy sectors.

**Table I.1 The three possible MROI variant A, B, and C.**

| Variant name | Main assumptions in methodology |
|---|---|
| A | $MROI_A = 1 + ((US.LTIR + 10)/100)$. |
| B | $MROI_B = 1 + AMEX\ Oil\ Index_{estimated}$. |
| C | $MROI_c = 1.1$ |

---

[1] The NYSE ARCA Oil Index, previously AMEX Oil Index, ticker symbol XOI, is a price-weighted index of the leading companies involved in the exploration, production, and development of petroleum. It measures the performance of the oil industry through changes in the sum of the prices of component stocks. The index was developed with a base level of 125 as of August 27th, 1984.

Regarding the variant B, the variable $AMEX\ Oil_{estimated}$ is computed following (I.1). Parameters values of (I.1) were obtained through a regression of the $AMEX\ Oil_{data}$ of Reuters (2016) on the $NYSE_{data}$ for the period 1984-2012.

$$AMEX\ Oil\ Index_{estimated} = 0.05466 + 0.65233 * NYSE_{data}. \tag{I.1}$$

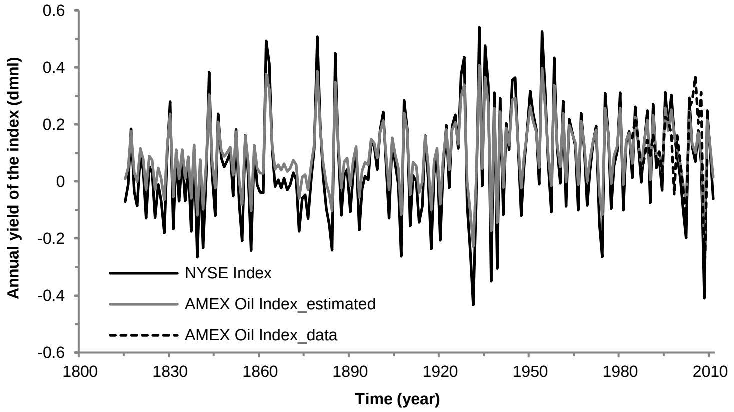


**Figure I.1 Reconstructed AMEX Oil Index annual yield, 1815–2012.**
**The AMEX Oil Index annual yield (grey line) is obtained with relation (I.1) where the NYSE Index data (black line) is retrieved from Goetzmann et al. (2001) for the 1815-1925 period, Ibbotson and Sinquefield (1976) for the 1926-1974 period, and NYSE (2015) for 1975-2012. The original AMEX Oil Index data (dashed grey line) of Reuters (2016) is only available for the period 1984-2012.**

Figure I.2 shows how the three MROI variants A, B, and C evolve over time.

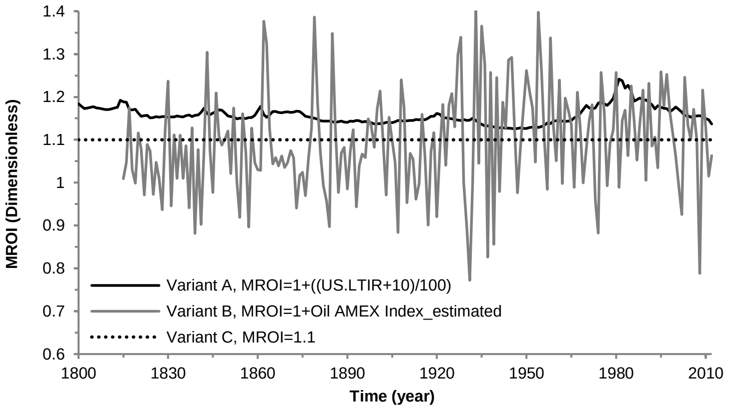


**Figure I.2 The three possible MROI variant A, B, and C**.

Figure I.3 presents our estimations of the global EROI of coal, oil, gas, and of the primary fossil energy system with the three possible MROI A, B, and C. It shows that our EROI results are quite insensitive to the MROI variability. Indeed, the three MROI variants deliver very consistent results. When looking at a particular energy type it is difficult to make a distinction between the different EROI estimations because methodological alternatives do not generates large enough output differences. This is particularly true for variant A and C which are hardly discernible. However, it is worth noting that there is a slightly higher volatility in values of variant B (that moreover cannot starts in 1800 because of the impossibility to estimate the $MROI_B$ before 1815).

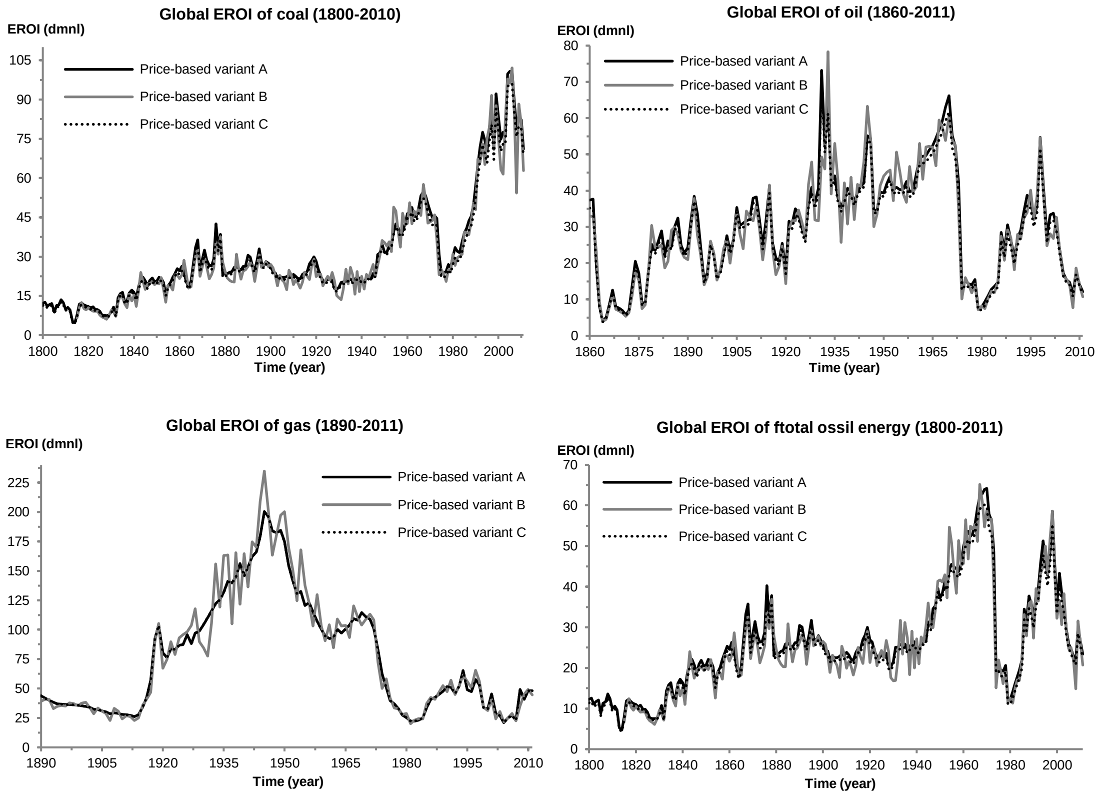


**Figure I.3 Impacts of variants A, B, and C on fossil fuels EROI.**

Sensitivity of price-based EROI estimates to the energy intensity

It is very likely that the different expenditures of the global fossil energy sector present overall a higher energy intensity than the average expenditures of the global economy. Indeed, the share of energy-intensive capital components such as steel is higher in the energy sector than in the global economy which relies on relatively more services (with less embodied energy). Thus, by taking the energy intensity of the global economy as a proxy for the energy intensity of the expenditures of the fossil energy sector, we should logically have overestimated the different EROI that we have calculated through our price-based methodology. This choice was made in order to have a time-dependent energy intensity, and indeed the energy intensity of the global economy has substantially decreased from 1800 (30 MJ/\$1990) to 2012 (10 MJ/\$1990). In their study concerning the EROI of US oil and gas production, Guilford et al. (2011) also used a national average of the energy intensity (8.3 MJ/\$2005, i.e. 12.4 MJ/\$1990), but they have then tested the sensitivity of their results with two other values: an estimate of the energy intensity of the American oil & gas industry expenditures of 14 MJ/\$2005 (i.e. 20.92 MJ/\$1990) based on the data of the Green Design Institute of Carnegie-Mellon University, and an arbitrary high estimate of 20 MJ/\$2005 (i.e. 29.9 MJ/\$1990). In Figure I.4 we show the effect of using 150% and 200% higher global energy intensities of expenditures on our price-based estimates of the global EROI of crude oil from 1860 to 2012. As previously anticipated, using the global energy intensity average tends to imply an overestimation of the resulting EROI. Nevertheless, the broad trend of crude oil global EROI is conserved and that is also true for coal and gas, so we can be confident in our main results: maximum global EROI seems to have been reached in the past for crude oil and gas, whereas increasing net energy yields are still to come for coal global production.

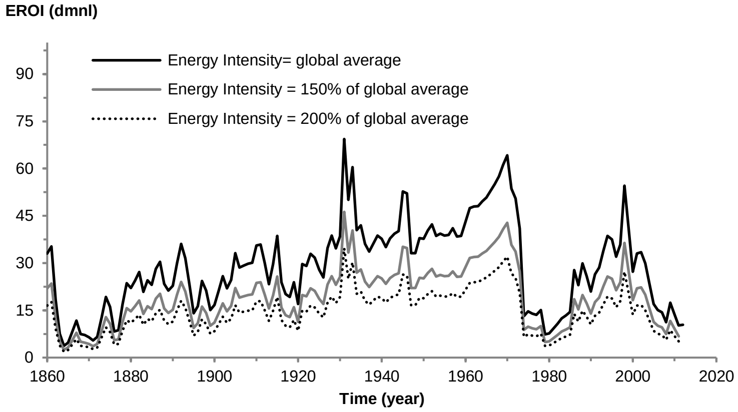


**Figure I.4 Sensitivity of the global EROI of crude oil to the energy intensity, 1860–2012.**

Comparison of results with existing studies

To check the robustness of our price-based estimations we use the work of Gagnon et al. (2009) in which an estimation of the global EROI of the combined oil and gas production is presented from 1992 to 2006. Hence, using the price-based method, we built an estimation of the global EROI of the joint oil and gas production and compared it to the one of Gagnon et al. (2009) in Figure I.5. Overall, all our variant estimations of the global EROI of oil and gas follows the same trend as the one of Gagnon et al. (2009): an increase between 1992 and 1999 followed by a decreasing phase up to 2006. Our estimation is globally higher and much more volatile than the one of Gagnon et al. (2009). This difference mostly comes from the irreducible volatility of energy prices we used, and the fact that we use a time-dependent energy intensity whereas in Gagnon et al. (2009) this variable is constant and worth 20 MJ/\$2005. To estimate the importance of the overall potential bias, we multiplied the denominator of the equation (4.24) by a parameter that we calibrated in order to minimize the sum of squared errors deriving from the difference between our estimation of the global EROI of oil and gas and the one of Gagnon et al. (2009) on the period 1992-2006. We found that on average our $EROI_A$ overestimate the one of Gagnon et al. (2009) by 20%. It is also worth noting that regarding the EROI of coal, values around 80:1 presented by our results in the last decade are perfectly in line with the estimation of the American coal EROI of Cleveland (2005).

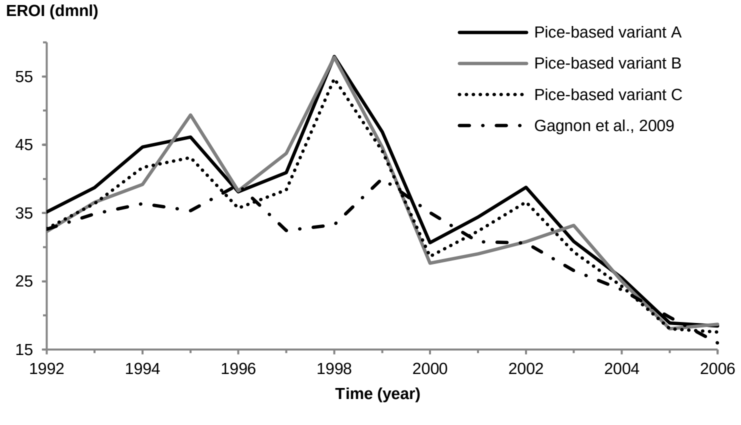


**Figure I.5 Gagnon et al. (2009) vs. price-based global EROI of oil and gas, 1992–2006.**

# APPENDIX J: ENERGY AND METAL REQUIREMENTS DATA

Data used in Section 5.1 and Section 5.2.

**Table J.1 Energy cost of metal productions and global metal productions in 2012.**

| Metal | Energy cost of production (GJ/tonne) | Minimum ore grade (%) | Source | Production (USGS, 2012) | Total energy cost (GJ) | Share of total energy consumption (%) |
|---|---|---|---|---|---|---|
| Aluminum | 212 | 17 | Rankin (2011) | 44400000 | 9391044000 | 1.798% |
| Antimony | 13 | 10 | Valero & Botero (2002) | 180000 | 2412000 | 0.000% |
| Arsenic | 28 | 2 | Valero & Botero (2002) | 46700 | 1307600 | 0.000% |
| Beryllium | 457.2 | 4 | Valero & Botero (2002) | 230 | 105156 | 0.000% |
| Bismuth | 56.4 | 5 | Valero & Botero (2002) | 7600 | 428640 | 0.000% |
| Cadmium | 110 | 0.4 | Valero & Botero (2002) | 21800 | 2398000 | 0.000% |
| Cerium | 354 | 20 | Tharumarajah & Koltun (2011) | 27000 | 9563400 | 0.002% |
| Chromium | 64 | 23 | Valero & Botero (2002) | 24000000 | 1538400000 | 0.295% |
| Cobalt | 322 | 0.2 | Valero & Botero (2002) | 110000 | 35420000 | 0.007% |
| Copper (hydro) | 64 | 0.5 | Rankin (2011) | 17000000 | 1095820000 | 0.210% |
| Copper (pyro) | 33 | 0.5 | Rankin (2011) | 17000000 | 561340000 | 0.107% |
| Gadolinium | 2162 |  | Tharumarajah & Koltun (2011) | 5000 | 10812000 | 0.002% |
| Gallium | 12660 | 0.01 | Valero & Botero (2002) | 200 | 2532000 | 0.000% |
| Germanium | 2215 | 6 | Valero & Botero (2002) | 118 | 261370 | 0.000% |
| Gold | 68400 | 0.0015 | Rankin (2011) | 2700 | 184680000 | 0.035% |
| Hafnium | 633 | 0.04 | Valero & Botero (2002) | 90 | 56970 | 0.000% |
| Indium | 2875 | 0.01 | Valero & Botero (2002) | 600 | 1725000 | 0.000% |
| Iridium | 2100 |  | Ashby (2012) | 4 | 8400 | 0.000% |
| Iron | 50 | 29 | Rankin (2011) | 1500000000 | 34050000000 | 6.519% |
| Lanthanum | 219 |  | Tharumarajah & Koltun (2011) | 25000 | 5485000 | 0.001% |
| Lead | 20 | 4 | Rankin (2011) | 5200000 | 101764000 | 0.019% |
| Lead (ISP) | 33 | 4 | Rankin (2011) | 5200000 | 169052000 | 0.032% |
| Lithium | 433 | 3 | Valero & Botero (2002) | 37000 | 16002500 | 0.003% |
| Magnesium | 437.3 | 0.1 | Valero & Botero (2002) | 6350000 | 2776855000 | 0.532% |
| Manganese | 56.9 | 25 | Valero & Botero (2002) | 17000000 | 967300000 | 0.185% |
| Mercury | 409 | 0.1 | Valero & Botero (2002) | 1810 | 740290 | 0.000% |
| Molybdenum | 148 | 0.3 | Valero & Botero (2002) | 250000 | 37000000 | 0.007% |
| Neodymium | 392 |  | Tharumarajah & Koltun (2011) | 21080 | 8263360 | 0.002% |
| Nickel (hydro) | 194 | 0.9 | Rankin (2011) | 2100000 | 406917000 | 0.078% |
| Nickel (pyro) | 114 | 0.9 | Rankin (2011) | 2100000 | 238392000 | 0.046% |
| Palladium | 5500 |  | Ashby (2012) | 200 | 1100000 | 0.000% |
| Platinum | 270500 | 0.00195 | Ashby (2012) | 179 | 48419500 | 0.009% |
| Praseodymium | 220 |  | Tharumarajah & Koltun (2011) | 2800 | 616280 | 0.000% |
| Rhenium | 171 | 0.3 | Valero & Botero (2002) | 5 | 855 | 0.000% |
| Rhodium | 14200 |  | Ashby (2012) | 25 | 355000 | 0.000% |
| Silver | 1582 | 0.01 | Valero & Botero (2002) | 24000 | 37968000 | 0.007% |

| | | | | | | |
|---|---|---|---|---|---|---|
| Tantalum | 1755 | 0.1 | Valero & Botero (2002) | 765 | 1342575 | 0.000% |
| Tin | 207 | 0.4 | Valero & Botero (2002) | 230000 | 47518000 | 0.009% |
| Titanium | 430 | 10 | Valero & Botero (2002) | 190000 | 81662000 | 0.016% |
| Tungsten | 357 | 0.6 | Valero & Botero (2002) | 75700 | 27024900 | 0.005% |
| Vanadium | 517 | 0.6 | Valero & Botero (2002) | 74000 | 38258000 | 0.007% |
| Yttrium | 756 | | Tharumarajah & Koltun (2011) | 10000 | 7559000 | 0.001% |
| Zinc (electrolytic) | 48 | 3.5 | Rankin (2011) | 13000000 | 629720000 | 0.121% |
| Zinc (ISP) | 36 | 3.5 | Rankin (2011) | 13000000 | 466050000 | 0.089% |
| Zirconium | 1371.5 | 2 | Valero & Botero (2002) | 1440000 | 1974960000 | 0.378% |
| **Metal sector energy consumption in 2012 (GJ):** | | | | | **52211442690** | **10.525%** |
| Primary energy production in 2012 (GJ): | | | | | 5.22345E+11 | 100.000% |

**Table J.2 Metal requirement per electricity producing technology.**

| Ton per MW | Parabolic trough | Solar tower plant | PV single si | PV multi si | PV a Si | PV CIGS | PV CdTe | Onshore wind power | Offshore wind power | Nuclear Power (PWR) | Hydro Power |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Cadmium | | | | | 0.00661 | 0.265 | 0.24426 | | | | |
| Chromium | 2.2 | 3.7 | | | 0.634 | | 0.061 | 0.3589 | 0.29356 | 0.35 | 1.5 |
| Copper | 3.2 | 1.4 | 0.825 | 0.943 | 1.005 | 0.45 | 5.1807 | 1.012 | 1.484 | 1.345 | 1.05 |
| Galium | | | | | | 0.124 | | | | | |
| Indium | | | | | 0.01341 | 0.055 | | | | | |
| Lead | | | 0.00553 | 0.00632 | | | 0.0078 | | | 0.04 | 0.3 |
| Molybdenum | 0.2 | 0.056 | | | 0.01062 | 0.109 | 0.0005 | 0.0753 | 0.075315 | | 0.25 |
| Nickel | 0.94 | 1.8 | 0.0013 | 0.0013 | 0.334 | | 0.01559 | 0.3766 | 0.37657 | 0.3 | |
| Niobium | | | | | | | | 0.0377 | 0.037657 | | |
| Selenium | | | | | | 0.11 | | | | | |
| Silver | 0.01342 | 0.01702 | 0.059 | 0.06817 | | | | | | | |
| Telurium | | | | | 0.0075 | | 0.24287 | | | | |
| Tin | | | | | 0.14321 | | | | | | |
| Vanadium | 0.0019 | 0.0017 | | | | | | 0.0904 | 0.0903787 | | |
| Zinc | 0.65 | 1.4 | 0.01562 | 0.01785 | | 0.121 | | | | | 0.4 |
| Praseodymium | | | | | | | | 0.0013 | 0.0308 | | |
| Neodymium | | | | | | | | 0.0062 | 0.15092 | | |
| Terbium | | | | | | | | 0.0003 | 0.00616 | | |
| Dysprosium | | | | | | | | 0.0009 | 0.02156 | | |

# APPENDIX K: UNIT ROOT TESTS FOR TIME SERIES USED IN SECTION 6.3

**Table K.1 Unit root tests for the different time series used in Section 6.3.**

| | **Augmented Dickey Fuller H0: Unit root** | | | **KPSS H0: Stationarity** | |
|---|---|---|---|---|---|
| | Constant +trend | Constant | None | Constant | Constant +trend |
| **1960-2010** | | | | | |
| US oil expenditure (as a fraction of GDP) | -1.822384 | -1.844912 | -0.870011 | 0.122781 | 0.125029* |
| US fossil expenditure (as a fraction of GDP) | -1.687235 | -1.710704 | -0.674999 | 0.126273 | 0.127640* |
| US total expenditure excluding wood (as a fraction of GDP) | -1.465321 | -1.514866 | -0.338669 | 0.144865 | 0.145643* |
| US total expenditure including wood (as a fraction of GDP) | -1.427043 | -1.463141 | -0.356135 | 0.144510 | 0.146239** |
| **1960-2010 + dummies for 1974 and 1979** | | | | | |
| US oil expenditure (as a fraction of GDP) | -4.645914*** | -4.569183*** | -3.968411*** | | |
| US fossil expenditure (as a fraction of GDP) | -3.855898** | -3.839113*** | -3.216946*** | | |
| US total expenditure excluding wood (as a fraction of GDP) | -3.391457* | -3.332861** | -2.655416*** | | |
| US total expenditure including wood (as a fraction of GDP) | -3.374901* | -3.349661** | -2.697435** | | |
| **1980-2010** | | | | | |
| US oil expenditure (as a fraction of GDP) | -2.544073 | -4.054355*** | -3.577915*** | 0.318747 | 0.185862** |
| US fossil expenditure (as a fraction of GDP) | -2.141382 | -3.517222** | -3.084142 | 0.339279 | 0.185632** |
| US total expenditure excluding wood (as a fraction of GDP) | -1.664036 | -3.305788** | -2.801680*** | 0.428181* | 0.185309** |
| US total expenditure including wood (as a fraction of GDP) | -1.691725 | -3.403323** | -2.912801*** | 0.448936* | 0.184505** |
| **1960-2010** | | | | | |
| US population | -0.491776 | 1.621706 | 18.19552 | 0.954076*** | 0.218951*** |
| US population first difference | -6.618667*** | -6.349020*** | -0.839929 | 0.383025* | 0.108030 |
| US unemployment rate | -2.987014 | -2.977318** | 0.019169 | 0.140524 | 0.124060* |
| US capital formation (as a fraction of GDP) | -2.784603 | -2.201140 | -0.637669 | 0.435677* | 0.070891 |
| US capital formation (as a fraction of GDP) + dummy in 2009 | -1.402460 | -3.106758** | -0.292106 | | |
| U S GDP growth rate | -5.761052*** | -5.535544*** | -3.757663*** | 0.284838 | 0.077606 |

Note: * Significant at 10% level, ** 5% level, ***1% level.

# REFERENCES

Abramovitz, M., 1989. *Thinking About Growth and Other Essays on Economic Growth and Welfare*, Cambridge, UK: Cambridge University Press.

Acemoglu, D., 1996. A Microfoundation for Social Increasing Returns in Human Capital Accumulation. *Quarterly Journal of Economics*, 111, pp.779–804.

Acemoglu, D., 2002a. Directed Technical Change. *Review of Economic Studies*, 69(4), pp.781–809.

Acemoglu, D., 2007. Equilibrium Bias of Technology. *Econometrica*, 75(5), pp.1371–1409.

Acemoglu, D., 2009. *Introduction to Modern Economic Growth*, Princeton, NJ: Princeton University Press.

Acemoglu, D., 2003. Labor- and Capital-Augmenting Technical Change. *Journal of the European Economic Association*, 1(1), pp.1–37.

Acemoglu, D., 2005. Politics and economics in weak and strong states. *Journal of Monetary Economics*, 52, pp.1199–1226.

Acemoglu, D., 2002b. Technical Change, Inequality and the Labor Market. *Journal of Economic Literature*, 40(1), pp.7–72.

Acemoglu, D. et al., 2012. The Environment and Directed Technical Change. *American Economic Review*, 102(1), pp.131–166.

Acemoglu, D., 1998. Why Do New Technologies Complement Skills? Directed Technical Change and Wage Inequality. *Quarterly Journal of Economics*, 113(4), pp.1055–1089.

Acemoglu, D. & Angrist, J., 2000. How Large are Human-Capital Externalities? Evidence from Compulsory-Schooling Laws. *NBER Macroeconomics Annual*, 15, pp.9–74.

Acemoglu, D., Antras, P. & Helpman, E., 2007. Contracts and Technology Adoption. *American Economic Review*, 97(3), pp.916–943.

Acemoglu, D. & Guerrieri, V., 2008. Capital Deepening and Nonbalanced Economic Growth. *Journal of Political Economy*, 116(3), pp.467–498.

Acemoglu, D., Johnson, S. & Robinson, J.A., 2005. Institutions as a Fundamental Cause of Long-Run Growth. In P. Aghion & S. Durlauf, eds. *Handbook of Economic Growth, Vol 1A*. Amsterdam, NL: Elsevier B.V., pp. 385–472.

Acemoglu, D., Johnson, S. & Robinson, J.A., 2002. Reversal of Fortune: Geography and Institutions in the Making of the Modern World Income Distribution. *Quarterly Journal of Economics*, 117(4), pp.1231–1294.

Acemoglu, D., Johnson, S. & Robinson, J.A., 2001. The Colonial Origins of Comparative Development: An Empirical Investigation. *American Economic Review*, 91(5), pp.1369–1401.

Acemoglu, D. & Robinson, J., 2010. The Role of Institutions in Growth and Development. *Review of Economics and Institutions*, 1(2), pp.1–30.

Acemoglu, D. & Robinson, J., 2012. *Why Nations Fail: The Origins of Power, Prosperity, and Poverty*, New York, NY: Crown Publishers.

Acemoglu, D. & Robinson, J.A., 2006. Economic Backwardness in Political Perspective. *American Political Science Review*, 100(1), pp.115–131.

Acemoglu, D. & Robinson, J.A., 2000a. Political Losers as a Barrier to Economic Development. *American Economic Review*, 90(2), pp.126–130.

Acemoglu, D. & Robinson, J.A., 2000b. Why Did the West Extend the Franchise? Democracy, Inequality, and Growth in Historical Perspective. *Quarterly Journal of Economics*, 115(4), pp.1167–1199.

Acemoglu, D. & Ventura, J., 2002. The World Income Distribution. *Quarterly Journal of Economics*, 117(2), pp.659–694.

Acemoglu, D. & Zilibotti, F., 1999. Information Accumulation in Development. *Journal of Economic Growth*, 4(1), pp.5–38.

Acemoglu, D. & Zilibotti, F., 2001. Productivity Differences. *Quarterly Journal of Economics*, 116(2), pp.563–606.

Acemoglu, D. & Zilibotti, F., 1997. Was Prometheus Unbound by Chance? Risk, Diversification, and Growth. *Journal of Political Economy*, 105(4), pp.709–751.

Adams, H., 1891. *History of the United States of America (1801-1817)*, New York, NY.

Aghion, P. & Howitt, P., 1992. A Model of Growth Through Creative Destruction. *Econometrica*, 60(2), pp.323–351.

Aghion, P. & Howitt, P., 1998. *Endogenous Growth Theory*, Cambridge, MA: MIT Press.

Aghion, P. & Howitt, P., 2009. *The Economics of Growth*, Cambridge, MA: MIT Press.

Aghion, P., Howitt, P. & Mayer-Foulkes, D., 2005. The Effect of Financial Development on Convergence: Theory and Evidence. *Quarterly Journal of Economics*, 120(1), pp.173–222.

Aiyar, S., Dalgaard, C.-J. & Moav, O., 2008. Technological progress and regress in pre-industrial times. *Journal of Economic Growth*, 13, pp.125–144.

Alesina, A. et al., 2003. Fractionalization. *Journal of Economic Growth*, 8, pp.155–194.

Alesina, A. & La Ferrara, E., 2005. Ethnic Diversity and Economic Performance. *Journal of Economic Literature*, 43(3), pp.762–800.

Alesina, A. & Fuchs-Schündeln, N., 2007. The Effect of Communism on People’s Preferences. *American Economic Review*, 97(4), pp.1507–1528.

Alesina, A. & Rodrik, D., 1994. Distributive Politics and Economic Growth. *Quarterly Journal of Economics*, 109(2), pp.465–490.

Allen, R.C., Bengtsson, T. & Dribe, M. eds., 2005. *Living Standards in the Past: New Perspectives on Well-Being in Asia and Europe*, Oxford, UK: Oxford University Press.

Alroy, J., 2001. A Multispecies Overkill Simulation of the End-Pleistocene Megafaunal Mass Extinction. *Science*, 292, pp.1893–1896.

Alston, L.J., Eggertsson, T. & North, D.C. eds., 1996. *Empirical Studies in Institutional Change*, Cambridge, UK: Cambridge University Press.

An, L., Jin, X. & Ren, X., 2014. Are the macroeconomic effects of oil price shock symmetric?: A Factor-Augmented Vector Autoregressive approach. *Energy Economics*, 45, pp.217–228.

Angrist, J., Lavy, V. & Schlosser, A., 2010. Multiple Experiments for the Causal Link between the Quantity and Quality of Children. *Journal of Labor EconomicsJournal of Labor Economics*, 28(4), pp.773–824.

Annila, A. & Salthe, S., 2009. Economies Evolve by Energy Dispersal. *Entropy*, 11(4), pp.606–633.

Arbatli, C.E., Ashraf, Q. & Galor, O., 2015. The Nature of Conflict. *NBER Working Paper*, pp.0–96.

Arcand, J.-L., Berkes, E. & Panizza, U., 2012. Too Much Finance? *IMF Working Paper*, 12(161).

Arcand, J.-L., Guillaumont, P. & Guillaumont Jeanneney, S., 2000. How to make a tragedy: on the alleged effect of ethnicity on growth. *Journal of International Development*, 12(7), pp.925–938.

Arrow, K.J., 1962. The Economic Implications of Learning by Doing. *Review of Economic*

*Studies*, 29(3), pp.155–173.

Arruñada, B., 2010. Protestants and Catholics: Similar Work Ethic, Different Social Ethic. *Economic Journal*, 120(547), pp.890–918.

Arvesen, A. & Hertwich, E.G., 2015. More caution is needed when using life cycle assessment to determine energy return on investment (EROI). *Energy Policy*, 76, pp.1–6.

Ashby, M.F., 2012. *Materials and the Environment Eco-Informed Material Choice,* Oxford, UK: Butterworth-Heinemann.

Ashraf, Q. & Galor, O., 2011. Dynamics and Stagnation in the Malthusian Epoch. *American Economic Review*, 101(5), pp.2003–2041.

Ashraf, Q. & Galor, O., 2013a. Genetic Diversity and the Origins of Cultural Fragmentation. *American Economic Review*, 103(3), pp.528–533.

Ashraf, Q. & Galor, O., 2014. Response to Comments made in a Letter by Guedes et al. on "The Out of Africa Hypothesis, Human Genetic Diversity and Comparative Development." *Open letter*.

Ashraf, Q. & Galor, O., 2016. The Macrogenoeconomics of Comparative Development. *Brown University Working Paper*, pp.0–40.

Ashraf, Q. & Galor, O., 2013b. The "Out of Africa" Hypothesis, Human Genetic Diversity, and Comparative Economic Developmenf. *American Economic Review*, 103(1), pp.1–46.

Ashraf, Q., Galor, O. & Klemp, M., 2015. Heterogeneity and Productivity. *Brown University Working Paper*, pp.0–36.

Atkins, P.W., 2010. *The Laws of Thermodynamics: A Very Short Introduction*, Oxford, UK: Oxford University Press.

Atkinson, A.B. & Stiglitz, J.E., 1969. A New View of Technological Change. *Economic Journal*, 79(315), pp.573–578.

Atlason, R. & Unnthorsson, R., 2014. Ideal EROI (energy return on investment) deepens the understanding of energy systems. *Energy*, 67, pp.241–245.

Autor, D.H., Katz, L.F. & Krueger, A.B., 1998. Computing Inequality: Have Computers Changed the Labor Market? *Quarterly Journal of Economics*, 113(4), pp.1169–1213.

Ayres, R.U., 1998. Eco-thermodynamics: economics and the second law. *Ecological Economics*, 26, pp.189–209.

Ayres, R.U., 2007. On the Practical Limits to Substitution. *Ecological Economics*, 61(1), pp.115–128.

Ayres, R.U., 1999. The Second Law, the Fourth Law, Recycling and Limits to Growth. *Ecological Economics*, 29(3), pp.473–483.

Ayres, R.U. et al., 2013. The underestimated contribution of energy to economic growth. *Structural Change and Economic Dynamics*, 27, pp.79–88.

Ayres, R.U. & Warr, B., 2009. *The Economic Growth Engine: How Energy and Work Drive Material Prosperity*, Cheltenham, UK: Edward Elgar Publishing.

Azariadis, C. & Drazen, A., 1990. Threshold Externalities in Economic Development. *Quarterly Journal of Economics*, 105, pp.501–526.

Banerjee, A. V. & Newman, A.F., 1998. Information, the Dual Economy, and Development. *Review of Economic Studies*, 65(4), pp.631–653.

Banfield, E.C., 1958. *The Moral Basis of a Backward Society*, New York, NY: Free Press.

Bardi, U., 2011. *The Limits to Growth Revisited*, New York, NY: Springer.

Barro, R.J., 1999. Determinants of Democracy. *Journal of Political Economy*, 107(S6), pp.S158–S183.

Barro, R.J., 1991. Economic Growth in a Cross Section of Countries. *Quarterly Journal of Economics*, 106(2), pp.407–443.

Barro, R.J. & McCleary, R.M., 2003. Religion and Economic Growth Across Countries. *American Sociological Review*, 68(5), pp.760–781.

Barro, R.J. & Sala-i-Martin, X., 2004. *Economic Growth* 2nd ed., Cambridge, MA: MIT Press.

Barro, R.J. & Sala-I-Martin, X., 1992. Convergence. *Journal of Political Economy*, 100(2), pp.223–251.

Bashmakov, I., 2007. Three laws of energy transitions. *Energy Policy*, 35(7), pp.3583–3594.

Basu, S. & Weil, D.N., 1998. Appropriate Technology and Growth. *Quarterly Journal of Economics*, 113(4), pp.1025–1054.

Baumol, W.J., 1967. Macroeconomics of Unbalanced Growth: The Anatomy of Urban Crisis. *American Economic Review*, 57(3), pp.415–426.

Beal, C.M. et al., 2012. Comprehensive Evaluation of Algal Biofuel Production: Experimental and Target Results. *Energies*, 5(12), pp.1943–1981.

Becker, G.S., 1965. A Theory of the Allocation of Time. *Economic Journal*, 75, pp.493–517.

Becker, G.S., 1981. *A Treatise on the Family*, Cambridge, MA: Harvard University Press.

Becker, G.S., 1960. An Economic Analysis of Fertility. In G. S. Becker, ed. *Demographic and Economic Change in Developed Countries*. Princeton, NJ: Princeton University Press.

Becker, S.O., Cinnirella, F. & Woessmann, L., 2010. The trade-off between fertility and education: evidence from before the demographic transition. *Journal of Economic Growth*, 15, pp.177–204.

Becker, S.O. & Woessmann, L., 2010. The Effect of Protestantism on Education Before the Industrialization: Evidence from 1816 Prussia. *Economics Letters*, 107(2), pp.224–228.

Becker, S.O. & Woessmann, L., 2009. Was Weber Wrong? A Human Capital Theory of Protestant Economic History. *Quarterly Journal of Economics*, 124(2), pp.531–596.

Bejan, A., 2016. *The Physics of Life: The Evolution of Everything*, New York, NY: St. Martin's Press, Macmillan Publishers.

Ben-Porath, Y., 1967. The Production of Human Capital and the Life Cycle of Earnings. *Journal of Political Economy*, 75(1), pp.352–365.

Benabou, R., 2000. Unequal Societies: Income Distribution and the Social Contract. *American Economic Review*, 90(1), pp.96–129.

Bencivenga, V.R. & Smith, B.D., 1991. Financial Intermediation and Endogenous Growth. *Review of Economic Studies*, 58, pp.195–209.

Bengtsson, T., Campbell, C. & Lee, J. eds., 2009. *Life Under Pressure: Mortality and Living Standards in Europe and Asia, 1700-1900*, Cambridge, MA: MIT Press.

Benhabib, J. & Spiegel, M.M., 2000. The Role of Financial Development in Growth and Investment. *Journal of Economic Growth*, 5(5), pp.341–360.

Benhmad, F., 2013. Dynamic cyclical comovements between oil prices and US GDP: A wavelet perspective. *Energy Policy*, 57, pp.141–151.

Benjamin, D.J. et al., 2012. The Promises and Pitfalls of Genoeconomics. *Annual Review of Economics*, 4(1), pp.627–662.

Bentham, J., 1789. *An Introduction to the Principles of Morals and Legislation*, London, UK: Payne, T. and Son.

Berndt, E.., 1978. Aggregate energy, efficiency and productivity measurement. *Annual Review Energy*, 3, pp.225–273.

Bihouix, P. & De Guillebon, B., 2010. *Quel Futur pour les Métaux ? - Raréfaction des Métaux : Un Nouveau Défi pour la Société*, Orsay, FR: EDP sciences.

Black, S.E., Devereux, P.J. & Salvanes, K.G., 2005. The More the Merrier ? The Effect of Family Size and Birth Order on Children's Education. *Quaterly Journal of Economics*, 120(2), pp.669–700.

Blanchard, O., 1997. The Medium Run. *Brookings Papers on Economic Activity*, 2, pp.89–158.

Blaug, M., 1985. *Economic Theory in Retrospect* 4th ed., Cambridge, UK: Cambridge University Press.

Bloom, D.E. & Sachs, J.D., 1998. Geography, Demography, and Economic Growth in Africa.

*Brookings Papers on Economic Activity*, 2, pp.207–295.

Blumler, M.A., 1992. *Seed Weight and Environment in Mediterranean-Type Grasslands in California and Israel*. Phd thesis, University of California, Berkeley.

Boltzmann, L., 1902. *Leçons sur la Théorie des Gaz*, Paris, FR: Gauthier-Villars.

Boserup, E., 1965. *The Conditions of Agricultural Growth: The Economics of Agrarian Change under Population Pressure*, Chicago, IL: Aldine Publishing.

Boucekkine, R., de la Croix, D. & Licandro, O., 2003. Early Mortality Declines at the Dawn of Modern Growth. *Scandinavian Journal of Economics*, 105(3), pp.401–418.

Bourguignon, F. & Morrisson, C., 2002. Inequality Among World Citizens: 1820-1992. *American Economic Review*, 92(4), pp.727–744.

Brandt, A.R., 2008. Converting Oil Shale to Liquid Fuels: Energy Inputs and Greenhouse Gas Emissions of the Shell in Situ Conversion Process. *Environmental Science & Technology*, 42(19), pp.7489–7495.

Brandt, A.R., 2009. Converting Oil Shale to Liquid Fuels with the Alberta Taciuk Processor: Energy Inputs and Greenhouse Gas Emissions. *Energy & Fuels*, 23(12), pp.6253–6258.

Brandt, A.R., 2011. Oil Depletion and the Energy Efficiency of Oil Production: The Case of California. *Sustainability*, 3(12), pp.1833–1854.

Brandt, A.R., Dale, M. & Barnhart, C.J., 2013. Calculating systems-scale energy efficiency and net energy returns: A bottom-up matrix-based approach. *Energy*, 62, pp.235–247.

Brandt, A.R., Englander, J. & Bharadwaj, S., 2013. The energy efficiency of oil sands extraction: Energy return ratios from 1970 to 2010. *Energy*, 55, pp.693–702.

Braudel, F., 1995. *A History of Civilizations*, New York, NY: Penguin Books.

Braudel, F., 1996. *The Mediterranean and the Mediterranean World in the Age of Philip II*, Oakland, CA: University of California Press.

British Petroleum, 2015. *Statistical Review of World Energy 2015*, Available at: http://www.bp.com/en/global/corporate/energy-economics/statistical-review-of-world-energy.html.

Brook, B.W. & Bowman, D.M.J.S., 2002. Explaining the Pleistocene Megafaunal Extinctions: Models, Chronologies, and Assumptions. *Proceedings of the National Academy of Sciences*, 99(23), pp.14624–14627.

Buera, F.J. et al., 2009. The Rise of the Service Economy. *NBER Working Paper*, 14822, pp.1–57.

Butzer, K.W., 1976. *Early Hydraulic Civilization in Egypt: A Study in Cultural Ecology*, Chicago, IL: The University of Chicago Press.

Cáceres-Delphiano, J., 2006. The Impacts of Family Size on Investment in Child Quality. *Journal of Human Resources*, 41(4), pp.738–754.

de Candolle, A., 1885. *Histoire des Sciences et des Savants Depuis Deux Siècles*, Genève-Bale: H. Georg.

Card, D., 1999. The Causal Effect of Education on Earnings. In O. Ashenfelter & D. Card, eds. *Handbook of Labor economics, Vol. 3A*. Amsterdam, NL: North-Holland, p. 1801–1863.

Caselli, F. et al., 2001. The U.S. Structural Transformation and Regional Convergence: A Reinterpretation. *Journal of Political Economy*, 109(3), pp.584–616.

Caselli, F. & Coleman, W.J., 2006. The World Technology Frontier. *American Economic Review*, 96(3), pp.499–522.

Cass, D., 1965. Optimum Growth in an Aggregative Model of Capital Accumulation. *Review of Economic Studies*, 32(3), pp.233–240.

Cavalli-Sforza, L.L. & Feldman, M.W., 1981. *Cultural Transmission and Evolution: A Quantitative Approach*, Princeton, NJ: Princeton University Press.

Cecchetti, S.G. & Kharroubi, E., 2015. Why does financial sector growth crowd out real economic growth? *BIS Working Papers*, 490.

Cervellati, M. & Sunde, U., 2005. Human Capital Formation, Life Expectancy, and the Process of Development. *American Economic Review*, 95(5), pp.1653–1672.

Chen, P.-Y., Chen, S.-T. & Chen, C.-C., 2012. Energy consumption and economic growth—New evidence from meta analysis. *Energy Policy*, 44, pp.245–255.

Chew, K.J., 2014. The future of oil: unconventional fossil fuels. *Philosophical Transactions of the Royal Society*, 372, pp.1–34.

Ciccone, A., Peri, G. & Antonio Ciccone, A., 2006. Identifying Human-Capital Externalities: Theory with Applications. *Review of Economic Studies*, 73(2), pp.381–412.

Clark, G., 2007. *A Farewell to Alms: A Brief Economic History of the World*, Princeton, NJ: Princeton University Press.

Clark, G., 2008. In defense of the Malthusian interpretation of history. *European Review of Economic History*, 12(2), pp.175–199.

Clark, G. & Cummins, N., 2016. The Child Quality-Quantity Tradeoff , England, 1750-1879: A Fundamental Component of the Economic Theory of Growth is Missing ? *CEPR Discussion Papers*, pp.1–44.

Clausius, R., 1867. *The Mechanical Theory of Heat with its Applications to the Steam-Engine and to the Physical Properties of Bodies*, London, UK: John van Voorst.

Cleveland, C., 2005. Net energy from the extraction of oil and gas in the United States. *Energy*, 30, pp.769–782.

Cleveland, C.J., 1999. Biophysical Economics: From Physiocracy to Ecological Economics and Industrial Ecology. In J. Gowdy & K. Mayumi, eds. *Bioeconomics and Sustainability: Essays in Honor of Nicholas Gerogescu-Roegen*. Cheltenham, UK: Edward Elgar Publishing, pp. 125–154.

Cleveland, C.J. et al., 1984. Energy and the U.S. economy: a biophysical perspective. *Science*, 225(4665), pp.890–897.

Cleveland, C.J., Kaufmann, R.K. & Stern, D.I., 2000. Aggregation and the role of energy in the economy. *Ecological Economics*, 32, pp.301–317.

Cole, H.S.D. et al. eds., 1973. *Models of Doom: A Critique of The limits to Growth*, New York, NY: Universe Books.

Collier, P., 2000. Ethnicity, Politics and Economic Performance. *Economics & Politics*, 12(3), pp.225–245.

Common, M. & Perring, C., 1992. Towards an ecological economics of sustainability. *Ecological Economics*, 6, pp.7–34.

Cook, E., 1976. Limits to exploitation of nonrenewable resources. *Science*, 191, pp.677–682.

Cook, E., 1971. The Flow of Energy in an Industrial Society. *Scientific American*, 225(3), pp.134–144.

Corning, P.A., 2002. Thermoeconomics: Beyond the Second Law. *Journal of Bioeconomics*, 4(1), pp.57–88.

Costanza, R., 2004. Value Theory and Energy. In C. J. Cleveland & R. U. Ayres, eds. *Encyclopedia of Energy, Vol. 6*. Boston, MA: Elsevier Academic Press.

Cottrell, W.F., 1955. *Energy and Society*, New York, NY: McGraw-Hill.

Cournède, B., Denk, O. & Hoeller, P., 2015. Finance and Inclusive Growth. *OECD Economic Policy Papers*, 14.

Craig, J.R., Vaughan, D.J. & Skinner, B.J., 2001. *Resources of the Earth* 3rd ed., Englewood Cliffs, NJ: Prentice Hall.

Craig, P.P., 2001. Energy Limits on Recycling. *Ecological Economics*, 36(3), pp.373–384.

La Croix, D. de & Michel, P., 2002. *A Theory of Economic Growth: Dynamics and Policy in Overlapping Generations*, Cambridge, UK: Cambridge University Press.

Crosby, A.W., 1986. *Ecological Imperialism: The Biological Expansion of Europe, 900-1900*, Cambridge, UK: Cambridge University Press.

Crowson, P., 2012. Some Observations on Copper Yields and Ore Grades. *Resources Policy*, 37(1), pp.59–72.

Dale, M., Krumdieck, S. & Bodger, P., 2011. A Dynamic Function for Energy Return on Investment. *Sustainability*, 3, pp.1972–1985.

Dale, M., Krumdieck, S. & Bodger, P., 2012. Global energy modelling — A biophysical approach ( GEMBA ) Part 2 : Methodology. *Ecological Economics*, 73, pp.158–167.

Dalgaard, C.J. & Strulik, H., 2015. The physiological foundations of the wealth of nations. *Journal of Economic Growth*, 20(1), pp.37–73.

Daly, H.E., 1985. The Circular Flow of Exchange Value and the Linear Throughput of Matter-Energy: A Case of Misplaced Concreteness. *Review of Social Economy New Dimensions of Social Economics (DECEMBER*, 43(3), pp.279–297.

Damodaran, A., 2015. Total Beta by Industry Sector. *Risk/Discount Rate*. Available at: http://pages.stern.nyu.edu/~adamodar/.

Dao, N.T. & Dávila, J., 2013. Can geography lock a society in stagnation? *Economics Letters*, 120, pp.442–446.

Dasgupta, P. & Heal, G., 1974. The Optimal Depletion of Exhaustible Resources. *Review of Economic Studies Symposium on the Economics of Exhaustible Resources*, 41, pp.3–28.

Denison, E.F., 1979. *Accounting for slower economic growth : the United States in the 1970s*, Washington, DC: The Brookings Institution.

Denison, E.F., 1985. *Trends in American Economic Growth, 1929-1982*, Washington D.C.: Brookings Institution Press.

Denison, E.F., 1967. *Why Growth Rates Differ: Postwar Experience in Nine Western Countries*, Washington, DC: Brookings Institution.

Diamond, J.M., 1997. *Guns, Germs, and Steel: The Fates of Human Societies*, New York, NY: W.W. Norton & Co.

Dincer, O.C. & Uslaner, E.., 2010. Trust and Growth. *Public Choice*, 142, pp.59–67.

Diwan, I. & Rodrik, D., 1991. Patents, appropriate technology, and North-South trade. *Journal of International Economics*, 30, pp.27–37.

Dixit, A.K. & Stiglitz, J.E., 1977. Monopolistic Competition and Optimum Product Diversity. *American Economic Review*, 67(3), pp.297–308.

Djankov, S. et al., 2003. The New Comparative Economics. *Journal of Comparative Economics*, 31(4), pp.595–619.

Doepke, M., 2004. Accounting for Fertility Decline During the Transition to Growth. *Journal of Economic Growth*, 9(3), pp.347–383.

Dollar, D., 1992. Outward-Oriented Developing Economies Really Do Grow More Rapidly: Evidence from 95 LDCs. *Economic Development and Cultural Change*, 40(3), pp.523–544.

Domar, E.D., 1946. Capital Expansion, Rate of Growth, and Employment. *Econometrica*, 14(2), pp.137–147.

Duflo, E., 2004. The Medium Run Effects of Educational Expansion: Evidence From a Large School Construction Program in Indonesia. *Journal of Development Economics*, 74, pp.163–197.

Duranton, G., 2004. The Economics of Production Systems: Segmentation and Skill-Biased Change. *European Economic Review*, 48, pp.307–336.

Dyke, J. & Kleidon, A., 2010. The Maximum Entropy Production Principle: Its Theoretical Foundations and Applications to the Earth System. *Entropy*, 12, pp.613–630.

Easterly, W., 2008. Institutions: Top Down or Bottom Up? *American Economic Review*, 98(2), pp.95–99.

Easterly, W. & Levine, R., 1997. Africa’s Growth Tragedy: Policies and Ethnic Divisions. *Quarterly Journal of Economics*, 112(4), pp.1203–1250.

Easterly, W. & Levine, R., 2003. Tropics, germs, and crops: How endowments influence economic development. *Journal of Monetary Economics*, 50(1), pp.3–39.

Elshkaki, A. & Graedel, T.E., 2013. Dynamic Analysis of the Global Metals Flows and Stocks in Electricity Generation Technologies. *Journal of Cleaner Production*, 59, pp.260–273.

Ember, C.R., Ember, M.R. & Peregrine, P.N., 2012. *Human Evolution and Culture: Highlights of Anthropology*. 7th ed., Harlow, UK: Pearson.

Etemad, B. & Luciani, J., 1991. *World Energy Production, 1800-1985. Production Mondiale d'Énergie, 1800-1985*, Genève, CH: Librairie Droz.

Faber, M., Niemes, H. & Stephan, G., 1987. *Entropy, Environment and Resources: An Essay in Physio-economics*, Berlin, DE: Springer.

Fagnart, J.F. & Germain, M., 2014. *Can the energy transition be smooth?*, Brussels, BE.

Farrell, A.E. et al., 2006. Ethanol can contribute to energy and environmental goals. *Science*, 311(576), pp.506–508.

Feldman, G.., 1964. *Foundations of Soviet Strategy for Economic Growth: Selected Soviet Essays, 1924-1930* N. Spulber, ed., Bloomington, IN: Indiana University Press.

Feldstein, M. & Horioka, C., 1980. Domestic Saving and International Capital Flows. *Economic Journal*, 90(358), pp.314–329.

Felipe, J. & McCombie, J.S.L., 2005. How Sound Are the Foundations of the Aggregate Production Function? *Eastern Economic Journal*, 31(3), pp.467–488.

Ferguson, N., 2011. *Civilization: the West and the Rest*, New York, NY: Penguin Press.

Fernandes, S.D. et al., 2007. Global Biofuel Use, 1850-2000. *Global Biogeochemical Cycles*, 21(2), pp.1–15.

Fernandez, R. & Rogerson, R., 1996. Income Distribution, Communities, and the Quality of Public Education. *Quarterly Journal of Economics*, 111(1), pp.135–164.

Fischer, D.H., 1989. *Albion's Seed: Four British Folkways in America*, New York, NY: Oxford University Press.

Fisher, F.M., 1969. The Existence of Aggregate Production Functions. *Econometrica*, 37(4), pp.553–577.

Flynn, D.O. & Giráldez, A. eds., 1997. *Metals and Monies in an Emerging Global Economy*, Brookfield, VT: Variorum.

Fondja Wandji, Y.D., 2013. Energy consumption and economic growth: Evidence from Cameroon. *Energy Policy*, 61, pp.1295–1304.

Fouquet, R., 2011. Divergences in Long Run Trends in the Prices of Energy and Energy Services. *Review of Environmental Economics and Policy*, 5(2), pp.196–218.

Fouquet, R., 2008. *Heat, Power and Light: Revolutions in Energy Services*, Northampton, MA: Edward Elgar Publishing.

Fouquet, R., 2014. Long run demand for energy services: income and price elasticities over 200 years. *Review of Environmental Economics and Policy*, 8(2), pp.186–207.

Fouquet, R., 2010. The slow search for solutions: Lessons from historical energy transitions by sector and service. *Energy Policy*, 38(11), pp.6586–6596.

Frank, A.G., 1998. *ReOrient : Global Economy in the Asian Age*, Berkeley, CA: University of California Press.

Frankel, J.A. & Romer, D., 1999. Does Trade Cause Growth? *American Economic Review*, 89(3), pp.379–399.

Freise, J., 2011. The EROI of Conventional Canadian Natural Gas Production. *Sustainability*, 3(12), pp.2080–2104.

Fukuyama, F., 2005. *Trust: The Social Virtues and the Creation of Prosperity*, New York, NY: Free Press.

Gagnon, N., Hall, C.A.S. & Brinker, L., 2009. A preliminary investigation of energy return on energy investment for global oil and gas production. *Energies*, 2(3), pp.490–503.

Gallup, J.L., Sachs, J.D. & Mellinger, A., 1999. Geography and Economic Development. *International Regional Science Review*, 22(2), pp.179–232.

Galor, O., 2005. From Stagnation to Growth: Unified Growth Theory. In P. Aghion & S. N. Durlauf, eds. *Handbook of Economic Growth, Vol. 1A*. Amsterdam, NL: Elsevier North-Holland.

Galor, O., 2011. *Unified Growth Theory,* Princeton, NJ: Princeton University Press.

Galor, O. & Michalopoulos, S., 2012. Evolution and the growth process: Natural selection of entrepreneurial traits. *Journal of Economic Theory*, 147, pp.759–780.

Galor, O. & Moav, O., 2002. Natural Selection and the Origin of Economic Growth. *Quarterly Journal of Economics*, 117(4), pp.1133–1191.

Galor, O., Moav, O. & Vollrath, D., 2009. Inequality in Landownership, the Emergence of Human-Capital Promoting Institutions, and the Great Divergence. *Review of Economic Studies*, 76(1), pp.143–179.

Galor, O. & Mountford, A., 2006. Trade and the Great dDivergence: The Family Connection. *American Economic Review*, 96(2), pp.299–303.

Galor, O. & Mountford, A., 2008. Trading Population for Productivity: Theory and Evidence. *Review of Economic Studies*, 75(4), pp.1143–1179.

Galor, O. & Weil, D.N., 2000. Population, Technology, and Growth: From Malthusian Stagnation to the Demographic Transition and Beyond. *American Economic Review*, 90(4), pp.806–828.

Gambetta, D., 1988. Can We Trust Trust? In D. Gambetta, ed. *Trust : Making and Breaking Cooperative Relations*. New York, NY: Blackwell Publishers.

Garrett, T.J., 2014. Long-run evolution of the global economy – Part 1: Physical basis. *Earth's Future*, 2, pp.127–151.

Garrett, T.J., 2015. Long-run evolution of the global economy – Part 2: Hindcasts of innovation and growth. *Earth System Dynamics*, 6, pp.673–688.

Georgescu-Roegen, N., 1971. *The Entropy Law and the Economic Process.*, Cambridge, MA: Harvard University Press.

Gerschenkron, A., 1962. *Economic Backwardness in Historical Perspective: A Book of Essays*, Cambridge, MA: Belknap Press of Harvard University Press.

Giraud, G. & Kahraman, Z., 2014. How Dependent is Growth from Primary Energy? The Dependency ratio of Energy in 33 Countries (1970-2011). *CES Working Papers*, pp.1–27.

Glaeser, E.L., Porta, R. La & Shleifer, A., 2004. Do Institutions Cause Growth? *Journal of Economic Growth*, 9, pp.271–303.

Goeller, H.E. & Weinberg, A.M., 1978. The Age of Substitutability. *American Economic Review*, 68(6), pp.1–11.

Goetzmann, W.N., Ibbotson, R.G. & Peng, L., 2001. A new historical database for the NYSE 1815 to 1925: Performance and predictability. *Journal of Financial Markets*, 4(1), pp.1–32.

Goldemberg, J., Coelho, S.T. & Guardabassi, P., 2008. The sustainability of ethanol production from sugarcane. *Energy Policy*, 36(6), pp.2086–2097.

Goldin, C. et al., 1998. The Origins of Technology-Skill Complementarity. *Quarterly Journal of Economics*, 113(3), pp.693–732.

Goldsmith, R.W., 1969. *Financial Structure and Development,* New Haven, CT: Yale University Press.

Goldstone, J.A., 2002. Efflorescences and Economic Growth in World History: Rethinking the "Rise of the West" and the Industrial Revolution. *Journal of World History,* 13(2), pp.323–389.

Goldstone, J.A., 1991. *Revolution and Rebellion in the Early Modern World,* Berkeley, CA: University of California Press.

Goldstone, J.A., 2009. *Why Europe? The Rise of the West in World History, 1500-1850*, Boston, MA: McGraw-Hill Higher Education.

Goodsell, W., 1937. Housing and the Birth Rate in Sweden. *American Sociological Review*, 2(6), pp.850–859.

Goody, J., 2004. *Capitalism and Modernity: The Great Debate*, Cambridge, MA: Polity.

Grandell, L., Hall, C.A.S. & Höök, M., 2011. Energy Return on Investment for Norwegian Oil and Gas from 1991 to 2008. *Sustainability*, 3(12), pp.2050–2070.

Granger, C.W.J., 1969. Investigating Causal Relations by Econometric Models and Cross-spectral Methods. *Econometrica*, 37(3), pp.424–438.

Grantham, G., 2008. Explaining the industrial transition: a non-Malthusian perspective. *European Review of Economic History*, 12(2), pp.155–165.

Grayson, D.K., 1991. Late Pleistocene Mammalian Extinctions in North America: Taxonomy, Chronology, and Explanations. *Journal of World Prehistory*, 5(3), pp.193–231.

Greenwood, J. & Jovanovic, 1990. Financial Development, Growth, and the Distribution of Income. *Journal of Political Economy*, 98(5).

Greif, A., 1994. Cultural Beliefs and Organization of Society: A Historical and Theoretical Reflection on Collectivist and Individualist Societies. *Journal of Political Economy*, 102(5), pp.912–950.

Griliches, Z., 1969. Capital-Skill Complementarity. *Review of Economics and Statistics*, 51(4), pp.465–468.

Grosjean, P., 2011. The Weight of History on European Cultural Integration: A Gravity Approach. *American Economic Review*, 101(3), pp.504–508.

Grossman, G.M. & Helpman, E., 1991a. *Innovation and Growth in the Global Economy*, Cambridge, MA: MIT Press.

Grossman, G.M. & Helpman, E., 1991b. Quality Ladders in the Theory of Growth. *Review of Economic Studies*, 58(1), pp.43–61.

Guedes, J. d'Alpoim et al., 2013. Is Poverty in Our Genes? A Critique of Ashraf and Galor &quot;The "Out of Africa" Hypothesis, Human Genetic Diversity, and Comparative Economic Development&quot;; American Economic Review (Forthcoming). *Current Anthropology Current Anthropology*, 54(1), pp.71–79.

Guilford, M.C. et al., 2011. A new long term assessment of energy return on investment (EROI) for U.S. oil and gas discovery and production. *Sustainability*, 3(10), pp.1866–1887.

Guiso, L., Sapienza, P. & Zingales, L., 2006. Does Culture Affect Economic Outcomes? *Journal of Economic Perspectives*, 20(2), pp.23–48.

Guiso, L., Sapienza, P. & Zingales, L., 2003. People's Opium? Religion and Economic Attitudes. *Journal of Monetary Economics*, 50(1), pp.225–282.

Guiso, L., Sapienza, P. & Zingales, L., 2004. The Role of Social Capital in Financial Development. *American Economic Review*, 94(3), pp.526–556.

Gunderson, L.H. & Holling, C.S., 2001. *Panarchy: Understanding Transformations in Human and Natural Systems*, Washington, DC: Island Press.

Habakkuk, H.J., 1962. *American and British Technology in the Nineteenth Century: Search for Labor- Saving Inventions*, Cambridge, UK: Cambridge University Press.

Hache, E. & Lantz, F., 2013. Speculative trading and oil price dynamic: A study of the WTI market. *Energy Economics*, 36, pp.334–340.

Hall, C.A.S., Balogh, S. & Murphy, D.J.R., 2009. What is the Minimum EROI that a Sustainable Society Must Have? *Energies*, 2(1), pp.25–47.

Hall, C.A.S., Cleveland, C.J. & Kaufmann, R., 1986. *Energy and Resource Quality: The Ecology of the Economic Process*, New York, NY: Wiley-Interscience.

Hall, C.A.S. & Day, J.W., 1990. *Ecosystem Modeling in Theory and Practice: An Introduction with Case Histories*, Boulder, CO: University Press of Colorado.

Hall, C.A.S. & Day, J.W., 2009. Revisiting the Limits to Growth After Peak Oil. *American Scientist*, 97(3), p.230.

Hall, C.A.S. & Klitgaard, K.A., 2012. *Energy and the Wealth of Nations: Understanding the Biophysical Economy*, New York, NY: Springer.

Hall, C.A.S., Lambert, J.G. & Balogh, S.B., 2014. EROI of different fuels and the implications for society. *Energy Policy*, 64, pp.141–152.

Hall, R.E. & Jones, C.I., 1999. Why Do Some Countries Produce So Much More Output Per Worker Than Others? *Quarterly Journal of Economics*, 114(1), pp.83–116.

Halloran, D., 2008. Appendix I, Wave Energy: Potential, EROI, and Social and Environmental Impacts. *The Oil Drum*. Available at: http://www.theoildrum.com/node/3949.

Hamilton, J.D., 1983. Oil and the Macroeconomy since World War II. *Journal of Political Economy*, 91(2), pp.228–248.

Hanabusa, K., 2009. Causality relationship between the price of oil and economic growth in Japan. *Energy Policy*, 37(5), pp.1953–1957.

Hanushek, E.A. & Woessmann, L., 2008. The Role of Cognitive Skills in Economic Development. *Journal of Economic Literature*, 46(3), pp.607–668.

Harmsen, J.H.M., Roes, A.L. & Patel, M.K., 2013. The impact of copper scarcity on the efficiency of 2050 global renewable energy scenarios. *Energy*, 50, pp.62–73.

Harrod, R.F., 1939. An Essay on Dynamic Theory. *Economic Journal*, 49(193), pp.14–33.

Hartley, P. et al., 2016. Energy Sector Innovation and Growth: An Optimal Energy Crisis. *Energy Journal*, 37(1), pp.233–258.

Hausmann, R., Hwang, J. & Rodrik, D., 2007. What you export matters. *Journal of Economic Growth*, 12, pp.1–25.

Hawking, S., 1988. *A Brief History of Time: From the Big Bang to Black Holes*, New York, NY: Bantam Books.

Hazan, M. & Zoabi, H., 2006. Does longevity cause growth? A theoretical critique. *Journal of Economic Growth*, 11, pp.363–376.

Helpman, E., 1998. *General Purpose Technologies and Economic Growth*, Cambridge, MA: MIT Press.

Herendeen, R.A., 2015. Connecting net energy with the price of energy and other goods and services. *Ecological Economics*, 109, pp.142–149.

Herrmann-Pillath, C., 2011. The evolutionary approach to entropy: Reconciling Georgescu-Roegen' s natural philosophy with the maximum entropy framework. *Ecological Economics*, 70(4), pp.606–616.

Heun, M.K. & de Wit, M., 2012. Energy return on (energy) invested (EROI), oil prices, and energy transitions. *Energy Policy*, 40, pp.147–158.

Hicks, J., 1932. *The Theory of Wages*, London, UK: Macmillan.

Hobson, J.M., 2004. *The Eastern Origins of Western Civilisation*, Cambridge, UK: Cambridge University Press.

Hoffman, P.T., 2015. *Why Did Europe Conquer the World?*, Princeton, NJ: Princeton University Press.

Holling, C.S., 2001. Understanding the Complexity of Economic, Ecological, and Social Systems. *Ecosystems*, 4(5), pp.390–405.

Honnery, D. & Moriarty, P., 2009. Estimating global hydrogen production from wind. *International Journal of Hydrogen Energy*, 34(2), pp.727–736.

Hoogwijk, M., de Vries, B. & Turkenburg, W., 2004. Assessment of the global and regional geographical, technical and economic potential of onshore wind energy. *Energy Economics*, 26(5), pp.889–919.

van Hoorn, A. & Maseland, R., 2013. Does a Protestant Work Ethic Exist? Evidence from the Well-Being Effect of Unemployment. *Journal of Economic Behavior & Organization*, 91, pp.1–12.

Howard, J.L. & Westby, R.M., 2013. *U.S. timber production, trade, consumption and price statistics 1965–2011*, Madison, WI.

Howitt, P., 2000. Endogenous Growth and Cross-Country Income Differences. *American Economic Review*, 90(4), pp.829–846.

Hu, Y. et al., 2013. Energy Return on Investment (EROI) of China’s conventional fossil fuels: Historical and future trends. *Energy*, 54, pp.352–364.

Huang, B.-N., Hwang, M.J. & Yang, C.W., 2008a. Causal relationship between energy consumption and GDP growth revisited: A dynamic panel data approach. *Ecological Economics*, 67(1), pp.41–54.

Huang, B.-N., Hwang, M.J. & Yang, C.W., 2008b. Does more energy consumption bolster economic growth? An application of the nonlinear threshold regression model. *Energy Policy*, 36(2), pp.755–767.

Hubbert, K., 1956. *Nuclear Energy and the Fossil Fuels*, San Antonio, Texas.

Huntington, S.P. & Rogers D. Spotswood Collection., 1996. *The Clash of Civilizations and the Remaking of World Order*, New York, NY: Simon & Schuster.

Ibbotson, R.G. & Sinquefield, R.A., 1976. Stocks, Bonds, Bills, and Inflation: Year-by-Year Historical Returns (1926-1974). *The Journal of Business*, 49(1), pp.11–47.

IIASA, 2012. *Global Energy Assessment - Toward a Sustainable Future*, Vienna, AT.

Inglehart, R. & Baker, W.E., 2000. Modernization, Cultural Change, and the Persistence of Traditional Values. *American Sociological Review*, 65(1), pp.19–51.

International Energy Agency, 2015. World Energy Statistics and Balances, online data service 2015 edition. Available at: http://www.iea.org/Sankey/.

Jacob, M.C., 1997. *Scientific Culture and the Making of the Industrial West*, Oxford, UK: Oxford University Press.

Jacob, M.C., 1988. *The Cultural Meaning of the Scientific Revolution*, Philadelphia, PA: Temple University Press.

Jacobs, J., 1969. *The Economy of Cities*, New York, NY: Random House.

James, J.A. & Skinner, J.S., 1985. The Resolution of the Labor-Scarcity Paradox. *Journal of Economic History*, 45(3), pp.513–540.

Jamil, F. & Ahmad, E., 2010. The relationship between electricity consumption, electricity prices and GDP in Pakistan. *Energy Policy*, 38(10), pp.6016–6025.

Jevons, W.S., 1871. *The Theory of Political Economy.*, London, UK: Macmillan & Co.

Johnston, L. & Williamson, S., 2016. What Was the US GDP Then? *MeasuringWorth.com*. Available at: https://www.measuringworth.com/usgdp/.

Jones, C.I., 1998. *Introduction to Economic Growth*, New York, NY: W.W. Norton.

Jones, C.I., 1997. On the Evolution of the World Income Distribution. *Journal of Economic Perspectives*, 11(3), pp.19–36.

Jones, C.I., 1995. R&D-Based Models of Economic Growth. *Journal of Political Economy*, 103(4), pp.759–784.

Jones, C.I., 2005. The Shape of Production Functions and the Direction of Technical Change. *Quarterly Journal of Economics*, 120(2), pp.517–549.

Jones, E.L., 1981. *The European Miracle: Environments, Economies, and Geopolitics in the History of Europe and Asia*, Cambridge, UK: Cambridge University Press.

Jones, R.W., 1971. A three-factor model in theory, trade and history. In J. Bhagwati et al., eds. *Trade, balance of payments and growth*. Amsterdam, NL.

Jorgenson, D.W. (Dale W., 1995. *Productivity*, Cambridge, MA: MIT Press.

Jorgenson, D.W., Gollop, F.M. & Fraumeni, B.M., 1987. *Productivity and U.S. Economic Growth*, Cambridge, MA: Harvard University Press.

Jorgenson, D.W. & Griliches, Z., 1967. The Explanation of Productivity Change. *Review of*

*Economic Studies*, 34(3), pp.249–283.

Jouvet, P.-A. & Schumacher, I., 2012. Learning-by-doing and the costs of a backstop for energy transition and sustainability. *Ecological Economics*, 73, pp.122–132.

Kåberger, T. & Månsson, B., 2001. Entropy and economic processes — physics perspectives. *Ecological Economics*, 36, pp.165–179.

Kalimeris, P., Richardson, C. & Bithas, K., 2014. A meta-analysis investigation of the direction of the energy-GDP causal relationship: implications for the growth-degrowth dialogue. *Journal of Cleaner Production*, 67, pp.1–13.

Kamarck, A.M., 1976. *The Tropics and Economic Development: A Provocative Inquiry into the Poverty of Nations*, Baltimore, MD: John Hopkins University Press.

Kander, A., Malanima, P. & Warde, P., 2013. *Power to the People: Energy in Europe Over the Last Five Centuries*, Princeton, NJ: Princeton University Press.

Katircioglu, S.T. et al., 2015. Oil price movements and macroeconomic performance: Evidence from twenty-six OECD countries. *Renewable and Sustainable Energy Reviews*, 44, pp.257–270.

Katz, L. & Autor, D., 1999. Changes in the Wage Structure and Earnings Inequality. In O. Ashenfelter & D. Card, eds. *Handbook of Labor Economics, Vol. 3A*. Amsterdam, NL: North-Holland.

Kaufmann, R.., 1994. The relation between marginal product and price in U.S. energy markets. *Energy Economics*, 16, pp.145–158.

Kaysen, C., 1972. The Computer that Printed Out W*O*L*F*. *Foreign Affairs*, 50(4), pp.660–668.

Kehoe, P.J. & Perri, F., 2002. International Business Cycles with Endogenous Incomplete Markets. *Econometrica*, 70(3), pp.907–928.

Kennedy, C., 1964. Induced Bias in Innovation and the Theory of Distribution. *Economic Journal*, 74(295), pp.541–547.

Keynes, J., 1936. *The General Theory of Employment, Interest and Money*, London, UK: Palgrave Macmillan.

King, C., 2015. Comparing World Economic and Net Energy Metrics, Part 3: Macroeconomic Historical and Future Perspectives. *Energies*, 8(12), pp.12997–13020.

King, C., Maxwell, J. & Donovan, A., 2015a. Comparing World Economic and Net Energy Metrics, Part 1: Single Technology and Commodity Perspective. *Energies*, 8(12), pp.12949–12974.

King, C., Maxwell, J. & Donovan, A., 2015b. Comparing World Economic and Net Energy Metrics, Part 2: Total Economy Expenditure Perspective. *Energies*, 8(12), pp.12975–12996.

King, C.W., 2014. Matrix method for comparing system and individual energy return ratios when considering an energy transition. *Energy*, 72, pp.254–265.

King, C.W. & Hall, C.A.S., 2011. Relating Financial and Energy Return on Investment. *Sustainability*, 3(10), pp.1810–1832.

Kleijn, R. et al., 2011. Metal Requirements of Low-Carbon Power Generation. *Energy*, 36(9), pp.5640–5648.

Klenow, P.J. & Rodriguez-Clare, A., 1997. The Neoclassical Revival in Growth Economics: Has It Gone Too Far? *NBER Macroeconomics Annual*, 12, pp.73–103.

Knack, S. & Keefer, P., 1997. Does Social Capital Have an Economic Payoff? A Cross-Country Investigation. *Quarterly Journal of Economics*, 112(4), pp.1251–1288.

Knack, S. & Keefer, P., 1995. Institutions and Economic Performance: Cross-country Tests Using Alternative Institutional Measures. *Economics & Politics*, 7(3), pp.207–227.

Kocaaslan, O.K., 2013. The causal link between energy and output growth: Evidence from Markov switching Granger causality. *Energy Policy*, 63, pp.1196–1206.

Kondratieff, N.D., 1935. The Long Waves in Economic Life. *Review of Economics and Statistics*, 17(6), pp.105–115.

Kongsamut, P., Rebelo, S. & Xie, D., 2001. Beyond Balanced Growth. *Review of Economic Studies*, 68(4), pp.869–882.

Koopmans, T.C., 1965. On the concept of optimal economic growth. *Pontificae Academiae Scientiarum Scripta Varia*, 28(1), pp.225–288.

Korotayev, A., Zinkina, J. & Bogevolnov, J., 2011. Kondratieff Waves in Global Invention Activity (1900-2008). *Technological Forecasting and Social Change*, 78(7), pp.1280–1284.

Korotayev, A.V. & Tsirel, S.V., 2010. A Spectral Analysis of World GDP Dynamics: Kondratieff Waves, Kuznets Swings, Juglar and Kitchin Cycles in Global Economic Development, and the 2008-2009 Economic Crisis. *Structure and Dynamics*, 4(1), pp.3–57.

Kraft, J. & Kraft, A., 1978. On the relationship between energy and GNP. *Journal of Energy and Development*, 3(2), pp.401–403.

Kreith, F. & Goswami, D.Y., 2007. *Handbook of Energy Efficiency and Renewable Energy*, Boca Raton, FL: Taylor & Francis Group, LLC.

Kreith, F. & Krumdieck, S., 2014. *Principles of Sustainable Energy Systems*, Boca Raton, FL: CRC Press, Taylor & Francis Group.

Kremer, M., 1993. Population Growth and Technological Change: One Million B.C. to 1990. *Quarterly Journal of Economics*, 108(3), pp.681–716.

Krugman, P., 1979. A Model of Innovation, Technology Transfer, and the World Distribution of Income. *Journal of Political Economy*, 87(2), pp.253–266.

Krumdieck, S. & Page, S., 2013. Retro-analysis of liquid bio-ethanol and bio-diesel in New Zeland. *Energy Policy*, 62, pp.363–371.

Krusell, P. et al., 2000. Capital-Skill Complementarity and Inequality: A Macroeconomic Analysis. *Econometrica*, 68(5), pp.1029–1053.

Krysiak, F.C., 2006. Entropy, limits to growth, and the prospects for weak sustainability. , 58, pp.182–191.

Kubiszewski, I., Cleveland, C.J. & Endres, P.K., 2010. Meta-analysis of net energy return for wind power systems. *Renewable Energy*, 35(1), pp.218–225.

Kümmel, R., 1989. Energy as a Factor of Production and Entropy as a Pollution Indicator in Macroeconomic Modelling. *Ecological Economics*, 1(2), pp.161–180.

Kümmel, R. et al., 1985. Technical progress and energy dependent production functions. *Zeitschrift für Nationalökonomie Journal of Economics*, 45(3), pp.285–311.

Kümmel, R., 1982. The impact of energy on industrial growth. *Energy*, 7(2), pp.189–203.

Kümmel, R., 2011. *The Second Law of Economics: Energy, Entropy, and the Origins of Wealth*, New York, NY: Springer.

Kümmel, R., Ayres, R.U. & Lindenberger, D., 2010. Thermodynamic Laws, Economic Methods and the Productive Power of Energy. *Journal of Non-Equilibrium Thermodynamics*, 35, pp.145–179.

Kümmel, R., Henn, J. & Lindenberger, D., 2002. Capital, labor, energy and creativity: modeling innovation diffusion. *Structural Change and Economic Dynamics*, 13(4), pp.415–433.

de la Croix, D. & Gosseries, A., 2012. The natalist bias of pollution control. *Journal of Environmental Economics and Management*, 63(2), pp.271–287.

Lagerlöf, N.-P., 2003. Gender Equality and Long-Run Growth. *Journal of Economic Growth*, 8, pp.403–426.

Lambert, J., Hall, C. & Balogh, S., 2013. *EROI of Global Energy Resources. Status, Trends and Social Implications*, Syracuse, USA.

Lambert, J.G. et al., 2014. Energy, EROI and quality of life. *Energy Policy*, 64, pp.153–167.

Landes, D.S., 1998. *The Wealth and Poverty of Nations: Why Some Are So Rich and Some So Poor*, New York, NY: W.W. Norton.

Lardic, S. & Mignon, V., 2008. Oil prices and economic activity: An asymmetric cointegration approach. *Energy Economics*, 30(3), pp.847–855.

Lebowitz, J.L., 2007. Entropy and phase transitions in macroscopic systems. *Physik Journal*, 6(8), pp.41–46.

Lee, J.Z. & Feng, W., 1999. *One Quarter of Humanity: Malthusian Mythology and Chinese Realities, 1700-2000*, Cambridge, MA: Harvard University Press.

Lenski, G., 1961. *The Religious Factor: A Sociological Study of Religion's Impact on Politics, Economics, and Family Life.*, Garden City, NY: Doubleday.

Leontief, W., 1953. Domestic Production and Foreign Trade; The American Capital Position Re-Examined. *Proceedings of the American Philosophical Society*, 97(4), pp.332–349.

Lewis, W.A., 1954. Economic Development with Unlimited Supplies of Labour. *The Manchester School*, 22(2), pp.139–191.

Li, H., Zhang, J. & Zhu, Y., 2008. The Quantity-Quality Trade-off of Children in a Developing Country: Identification Using Chinese Twins. *Demography*, 45(1), pp.223–243.

Lindenberger, D. & Kümmel, R., 2011. Energy and the state of nations. *Energy*, 36(10), pp.6010–6018.

Lipset, S., 1960. *Political Man: The Social Bases of Politics.*, Garden City, NY: Doubleday.

Lipsey, R.G., Carlaw, K. & Bekar, C., 2005. *Economic Transformations: General Purpose Technologies and Long-Term Economic Growth*, Oxford, UK: Oxford University Press.

Liu, H. et al., 2006. A Geographically Explicit Genetic Model of Worldwide Human-Settlement History. *American journal of human genetics*, 79(2), pp.230–237.

Loayza, N. V. et al., 2006. Financial Development, Financial Fragility, and Growth. *Journal of Money, Credit, and Banking*, 38(4).

Lotka, A.J., 1922a. Contribution to the Energetics of Evolution. *Proceedings of the National Academy of Sciences*, 8, pp.147–151.

Lotka, A.J., 1925. *Elements of Physical Biology*, Baltimore, MD: Williams & Wilkins Company.

Lotka, A.J., 1922b. Natural Selection as a Physical Principle. *Proceedings of the National Academy of Sciences*, 8, pp.151–154.

Lucas, R.E., 1988. On the Mechanics of Economic Development. *Journal of Monetary Economics*, 22, pp.3–42.

Lucas, R.E., 1990. Why Doesn't Capital Flow from Rich to Poor Countries? *American Economic Review*, 80(2), pp.92–96.

Lund, P.D., 2007. Upfront Resource Requirements for Large-Scale Exploitation Schemes of New Renewable Technologies. *Renewable Energy*, 32(3), pp.442–458.

Machiavelli, N., 1997. *Discourses on Livy*, New York, NY: Oxford University Press.

Maddison, A., 2007. *Contours of the World Economy, 1− 2030 AD: Essays in Macro-Economic History*, Oxford, UK: Oxford University Press.

Mahalanobis, P.C., 1953. Some Oberservations on the Process of Growth of National Income. *Sankhya: The Indian Journal of Statistics (1933-1960)*, 12(4).

Malthus, T.R., 1798. *An Essay on the Principle of Population, as it Affects the Future Improvement of Society with Remarks on the Speculations of Mr. Godwin, M. Condorcet, and Other Writers*, London, UK: J. Johnson.

Mankiw, G.N., Romer, D. & Weil, D.N., 1992. A Contribution to the Empirics of Economics Growth. *Quarterly Journal of Economics*, 107(2), pp.407–437.

Manthy, R.S., 1978. *Natural Resource Commodities - A Century of Statistics*, Washington, DC: John Hopkins Press for Resources for the Future.

Marks, R., 2002. *The Origins of the Modern World: A Global and Ecological Narrative*, Lanham, MD: Rowman & Littlefield.

Marshall, A., 1890. *Principles of Economics*, London, UK: Macmillan Publishers Ltd.

Martin, P., 1967. Prehistoric Overkill. In H. E. J. Wright & P. S. Martin, eds. *Pleistocene Extinctions: The Search for a Cause*. New Haven, CT: Yale University Press, pp. 75–120.

Martin, P.S., 1984. Prehistoric Overkill: The Global Model. In P. S. Martin & R. G. KLein, eds. *Quaternary Extinctions: A Prehistoric Revolution*. Tucson, AZ: University of Arizona Press, pp. 354–403.

Martyushev, L.M. & Seleznev, V.D., 2006. Maximum entropy production principle in physics, chemistry and biology. *Physics Reports*, 426(1), pp.1–45.

Martyushev, L.M. & Seleznev, V.D., 2014. The restrictions of the maximum entropy production principle. *Physica A: Statistical Mechanics and its Applications*, 410, pp.17–21.

Marx, K., 1867. *Das Kapital, Kritik der politischen Ökonomie*, Hamburg, DE: Verlag von Otto Meisner.

Maslow, A., 1943. A theory of human motivation. *Psychological Review*, 50, pp.370–396.

Matsuyama, K., 1992. Agricultural Productivity, Comparative Advantage, and Economic Growth. *Journal of Economic Theory*, 58, pp.317–334.

Matsuyama, K., 2004. Financial Market Globalization, Symmetry-Breaking and Endogenous Inequality of Nations. *Econometrica*, 72(3), pp.853–884.

Matsuyama, K., 2002. The Rise of Mass Consumption Societies. *Journal of Political Economy*, 110(5), pp.1035–1070.

McCloskey, D.N., 2008. “You know, Ernest, the rich are different from you and me”: a comment on Clark’s A Farewell to Alms. *European Review of Economic History*, 12(2), pp.138–148.

Mcglade, C. & Ekins, P., 2015. The geographical distribution of fossil fuels unused when limiting global warming to 2 °C. *Nature*, 517, pp.187–190.

McNeill, W., 1963. *The Rise of the West : A History of the Human Community*, Chicago, IL: University of Chicago Press.

Meadows, D.H. et al., 1972. *The Limits to Growth: A Report for the Club of Rome’s Project on the Predicament of Mankind*, New York, NY: Universe Books.

Memary, R. et al., 2012. Life cycle assessment: a time-series analysis of copper. *Journal of Cleaner Production*, 33, pp.97–108.

Menger, C., 1871. *Grundsätze der Volkswirtschaftslehre*, Vienna, AT: Braumüller.

Merton, R.K., 1938. Science, Technology and Society in Seventeenth Century England. *Osiris*, 4, pp.360–632.

Messner, F., 2002. Material substitution and path dependence: empirical evidence on the substitution of copper for aluminum. *Ecological Economics*, 42(1), pp.259–271.

Milanovic, B., 2011. A Short History of Global Inequality: The Past Two Centuries. *Explorations in Economic History*, 48(4), pp.494–506.

Milanovic, B., 2012. Global Inequality Recalculated and Updated: The Effect of New PPP Estimates on Global Inequality and 2005 Estimates. *Journal of Economic Inequality*, 10(1), pp.1–18.

Mill, J.S., 1863. *Utilitarianism*, London, UK: Parker, Son & Bourn.

Mincer, J., 1974. *Schooling, Experience, and Earnings.*, New York, NY: National Bureau of Economic Research; distributed by Columbia University Press.

Modahl, I.S. et al., 2013. How methodological issues affect the energy indicator results for different electricity generation technologies. *Energy Policy*, 63, pp.283–299.

Moerschbaecher, M. & Day, J.W., 2011. Ultra-Deepwater Gulf of Mexico Oil and Gas: Energy Return on Financial Investment and a Preliminary Assessment of Energy Return on Energy Investment. *Sustainability*, 3(12), pp.2009–2026.

Mohr, S.H. & Evans, G.M., 2009. Forecasting coal production until 2100. *Fuel*, 88(11), pp.2059–2067.

Mokyr, J., 1993. *The British Industrial Revolution: An Economic Perspective*, Boulder, CO: Westview Press.

Mokyr, J., 2009. *The Enlightened Economy: An Economic History of Britain, 1700-1850*, New Haven, CT: Yale University Press.

Mokyr, J., 2005. *The Gifts of Athena: Historical Origins of the Knowledge Economy*, Princeton, NJ: Princeton University Press.

Mokyr, J., 1990. *The Lever of Riches : Technological Creativity and Economic Progress*, New York, NY: Oxford University Press.

Montesquieu, C.L. de S. baron de L.B. et de, 1748. *De l'Esprit des Lois*, Genève, CH: Barrillot & Fils.

Moriarty, P. & Honnery, D., 2012. What is the global potential for renewable energy? *Renewable and Sustainable Energy Reviews*, 16, pp.244–252.

Morris, I., 2015. *Foragers, Farmers, and Fossil Fuels: How Human Values Evolve*, Princeton, NJ: Princeton University Press.

Morris, I., 2010. *Why the West Rules−For Now: The Patterns of History, and What They Reveal About the Future*, New York, NY: Farrar, Straus and Giroux.

Moss, R.L. et al., 2013. The Potential Risks from Metals Bottlenecks to the Deployment of Strategic Energy Technologies. *Energy Policy*, 55, pp.556–564.

Mudd, G.M. et al., 2013. A Detailed Assessment of Global Cu Resource Trends and Endowments. *Economic Geology*, 108(5), pp.1163–1183.

Mudd, G.M., 2010. The Environmental sustainability of mining in Australia: key mega-trends and looming constraints. *Resources Policy*, 35(2), pp.98–115.

Mudd, G.M. & Diesendorf, M., 2008. Sustainability of Uranium Mining and Milling: Toward Quantifying Resources and Eco-Efficiency. *Environmental Science & Technology*, 42(7), pp.2624–2630.

Murphy, D.J. et al., 2011. Order from Chaos: A Preliminary Protocol for Determining the EROI of Fuels. *Sustainability*, 3(10), pp.1888–1907.

Murphy, D.J. & Hall, C.A.S., 2011a. Adjusting the economy to the new energy realities of the second half of the age of oil. *Ecological Modelling*, 223(1), pp.67–71.

Murphy, D.J. & Hall, C.A.S., 2011b. Energy return on investment, peak oil, and the end of economic growth. *Annals of the New York Academy of Sciences*, 1219(1), pp.52–72.

Murphy, D.J. & Hall, C.A.S., 2010. Year in review-EROI or energy return on (energy) invested. *Annals of the New York Academy of Sciences*, 1185(1), pp.102–118.

Naccache, T., 2010. Slow oil shocks and the "weakening of the oil price–macroeconomy relationship." *Energy Policy*, 38(5), pp.2340–2345.

Neher, P.A., 1971. Peasants, Procreation, and Pensions. *American Economic Review*, 61(3), pp.380–389.

Nelson, R.R. & Phelps, E.S., 1966. Investment in Humans,Technological Diffusion, and Economic Growth. *American Economic Review*, 56(2), pp.69–75.

Nelson, R.R. & Winter, S.G., 1982. *An Evolutionary Theory of Economic Change*, Cambridge, MA: Harvard University Press.

von Neumann, J., 1945. A Model of General Economic Equilibrium. *Review of Economic Studies*, 13(1), pp.1–9.

Ngai, L.R. & Pissarides, C.A., 2007. Structural Change in a Multisector Model of Growth. *American Economic Review*, 97(1), pp.429–443.

Nordhaus, W., 2011. Designing a friendly space for technological change to slow global warming. *Energy Economics*, 33(4), pp.665–673.

Norgate, T. & Haque, N., 2010. Energy and greenhouse gas impacts of mining and mineral processing operations. *Journal of Cleaner Production*, 18(3), pp.266–274.

Norgate, T. & Jahanshahi, S., 2010. Low Grade Ores – Smelt, Leach or Concentrate? *Minerals Engineering*, 23(2), pp.65–73.

Norgate, T. & Jahanshahi, S., 2011. Reducing the Greenhouse Gas Footprint of Primary Metal Production: Where Should the Focus Be? *Minerals Engineering*, 24(14), pp.1563–1570.

North, D.C., 1994. Economic Performance Through Time. *American Economic Review*, 84(3), pp.359–368.

North, D.C., 1990. *Institutions, Institutional Change, and Economic Performance*, Cambridge, UK: Cambridge University Press.

North, D.C., 1981. *Structure and Change in Economic History*, New York, NY: W.W. Norton.

North, D.C. & Thomas, R.P., 1973. *The Rise of the Western World: A New Economic History*, Cambridge, UK: Cambridge University Press.

Northey, S., Haque, N. & Mudd, G., 2013. Using sustainability reporting to assess the environmental footprint of copper mining. *Journal of Cleaner Production*, 40, pp.118–128.

Nowak, R., 1991. *Walker's Mammals of the World*, Baltimore, MD: John Hopkins University Press.

NREL, 2010a. National Solar Radiation Database. Available at: http://rredc.nrel.gov/solar.

NREL, 2010b. Western Wind Dataset. Available at: http://www.nrel.gov/wind/integrationdatasets/western/data.html.

Nunn, N. & Wantchekon, L., 2011. The Slave Trade and the Origins of Mistrust in Africa. *American Economic Review*, 101(7), pp.3221–3252.

Nunziata, L. & Rocco, L., 2016. A Tale of Minorities: Evidence on Religious Ethics and Entrepreneurship. *Journal of Economic Growth*, 21(2), pp.189–224.

NYSE, 2015. NYSE Composite Index: Annual Returns. Available at: http://www.forecast-chart.com/historical-nyse-composite.html.

O'Connor, M., 1991. Entropy, structure, and organisational change. *Ecological Economics*, 3, pp.95–122.

Obstfeld, M., 1994. Risk-Taking, Global Diversification, and Growth. *American Economic Review*, 84(5), pp.1310–1329.

Obstfeld, M. & Taylor, A.M., 2003. Globalization and Capital Markets. In M. D. Bordo, A. M. Taylor, & J. G. Williamson, eds. *Globalization in Historical Perspective*. University of Chicago Press.

Odum, H.T., 1973. Energy, Ecology, and Economics. *AMBIO*, 2(6), pp.220–227.

Odum, H.T., 1971. *Environment, Power, and Society*, New York, NY: Wiley-Interscience.

Odum, H.T., 1995. Self-organization and Maximum Empower. In *Maximum power: The Ideas and Applications of H.T. Odum*. Boulder, CO: Colorado University Press.

Officer, L. & Williamson, S., 2016. The Annual Consumer Price Index for the United States, 1774-2015. *MeasuringWorth.com*. Available at: http://www.measuringworth.com/uscpi/.

Officer, L.H., 2016. What Was the Interest Rate Then? *MeasuringWorth.com*. Available at: http://www.measuringworth.com/interestrates/.

Ohlin, B., 1967. *Interregional and International Trade* 2nd ed., Cambridge, MA: Harvard University Press.

Olsson, O. & Hibbs, D.A., 2005. Biogeography and Long-Run Economic Development. *European Economic Review*, 49(4), pp.909–938.

Omri, A., 2014. An international literature survey on energy-economic growth nexus: Evidence from country-specific studies. *Renewable and Sustainable Energy Reviews*, 38, pp.951–959.

Ostwald, W., 1911. Efficiency. *The Independent*, 71, pp.867–871.

Overton, M., 1996. *Agricultural Revolution in England: the Transformation of the Agrarian*

*Economy, 1500-1850*, Cambridge, UK: Cambridge University Press.

Ozawa, H. et al., 2003. The Second Law of Thermodynamics and the Global Climate System: A Review of the Maximum Entropy Production Principle. *Reviews of Geophysics*, 41(4), pp.1–24.

Parente, S.L. & Prescott, E.C., 1994. Barriers to Technology Adoption and Development. *Journal of Political Economy*, 102(2), pp.298–321.

Perdue, P.C., 2005. *China marches west : the Qing conquest of Central Eurasia*, Cambridge, MA: Harvard University Press.

Persson, K.G., 2008. The Malthus delusion. *European Review of Economic History*, 12(2), pp.165–173.

Persson, T. & Tabellini, G., 1994. Is Inequality Harmful for Growth? *American Economic Review*, 84(3), pp.600–621.

Pigou, A., 1920. *The Economics of Welfare*, London, UK: Macmillan & Co.

Pimentel, D. & Patzek, T.W., 2005. Ethanol Production Using Corn, Switchgrass, and Wood; Biodiesel Production Using Soybean and Sunflower. *Natural Resources Research*, 14(1).

Planck, M., 1903. *Treatise on Thermodynamics*, London, UK: Longmans, Green & Co.

Van Der Ploeg, F., 2011. Natural Resources: Curse or Blessing? *Journal of Economic Literature*, 49(2), pp.366–420.

Podolinsky, S., 1880. Le Socialisme et l’Unité des Forces Physiques. *La Revue Socialiste*, 8. Available at: http://gallica.bnf.fr/ark:/12148/bpt6k58174297.r=Revue Socialiste 1880.langEN.

Poisson, A. & Hall, C., 2013. Time Series EROI for Canadian Oil and Gas. *Energies*, 6(11), pp.5940–5959.

Pomeranz, K., 2000. *The Great Divergence: China, Europe, and the Making of the Modern World Economy*, Princeton, NJ: Princeton University Press.

Popp, D., 2002. Induced Innovation and Energy Prices. *American Economic Review*, 92(1), pp.160–180.

Pradhan, A. et al., 2011. Energy Lige-Cycle Assessment of Soybean Biodiesel Revisited. *Transactions of the ASABE*, 54(3), pp.1031–1039.

Prados de la Escosura, L., 2015. Mismeasuring Long-Run Growth: The Bias from Splicing National Accounts—The Case of Spain. *Cliometrica*, online, pp.1–25.

Pressman, S., 2005. What Is Wrong with the Aggregate Production Function? *Eastern Economic Journal*, 31(3), pp.422–425.

Prigogine, I., 1967. Dissipative Strucutre in Chemical Systems. In S. Claesson, ed. *Fast Reactions and Primary Processes in Chemical Kinetics*. New York, NY: Interscience.

Przeworski, A. et al., 2000. *Democracy and Development: Political Institutions and Well-Being in the World, 1950-1990*, Cambridge, UK: Cambridge University Press.

Putnam, 1993. *Making Democracy Work*, Princeton, NJ: Princeton University Press.

Quigley, C., 1961. *The Evolution of Civilizations: An Introduction to Historical Analysis*, New York, NY: Macmillan & Co.

Raine, A., Foster, J. & Potts, J., 2006. The new entropy law and the economic process. *Ecological Complexity*, 3(4), pp.354–360.

Ramirez, P. & Hall, C.A.., 2013. The relation of oil to the Mexican economy: past, present and future. In E. Yáñez-Avancibia et al., eds. *Ecological dimensions for sustainable socioeconomic development*. WIT Press, p. 628.

Ramsey, F.P., 1928. A Mathematical Theory of Saving. *Economic Journal*, 38(152), pp.543–559.

Rankin, W.J., 2011. *Minerals, Metals and Sustainability: Meeting Future Material Needs*, Boca Raton, FL: CSIRO Publishing.

Raugei, M., Fullana-i-Palmer, P. & Fthenakis, V., 2012. The energy return on energy investment (EROI) of photovoltaics: Methodology and comparisons with fossil fuel life cycles. *Energy Policy*, 45, pp.576–582.

Rebelo, S., 1991. Long-Run Policy Analysis and Long-Run Growth. *Journal of Political Economy*, 99(3), pp.500–521.

Reuters, 2016. Amex Oil Index. Available at: http://www.reuters.com/finance/markets/index?symbol=.XOI.

Ricardo, D., 1817. *On the Principles of Political Economy and Taxation*, London, UK: John Murray.

Richerson, P.J. & Boyd, R., 2005. *Not by Genes Alone: How Culture Transformed Human Evolution*, Chicago, IL: University of Chicago Press.

Rodrik, D., Subramanian, A. & Trebbi, F., 2004. Institutions Rule: The Primacy of Institutions Over Geography and Integration in Economic Development. *Journal of Economic Growth*, 9, pp.131–165.

Romer, P.M., 1990. Endogenous Technological Change. *Journal of Political Economy*, 98(5), pp.71–102.

Romer, P.M., 1987. Growth Based on Increasing Returns Due to Specialization. *American Economic Review*, 77(2), pp.56–62.

Romer, P.M., 1986. Increasing Returns and Long-Run Growth. *Journal of Political Economy*, 94(5), pp.1002–1037.

Rostow, W.W., 1963. *The Economics of Take-Off Into Sustained Growth*, New York, NY: Macmillan Publishers Ltd.

Rühl, C. et al., 2012. Economic development and the demand for energy: A historical perspective on the next 20 years. *Energy Policy*, 50, pp.109–116.

Ruth, M., 1995. Information, order and knowledge in economic and ecological systems: implications for material and energy use. *Ecological Economics*, 13(2), pp.99–114.

Sachs, J.D. et al., 1995. Economic Reform and the Process of Global Integration. *Brookings Papers on Economic Activity*, 1(1), pp.1–118.

Saint-Paul, G. & Verdier, T., 1993. Education, democracy and growth. *Journal of Development Economics*, 42, pp.399–407.

Samuelson, P.A., 1971. Ohlin Was Right. *Swedish Journal of Economics*, 73(4), p.365_384.

Samuelson, P.A. & Solow, R.M., 1956. A Complete Capital Model involving Heterogeneous Capital Goods. *The Quarterly Journal of Economics*, 70(4), pp.537–562.

Samuelsson, K., 1961. *Religion and Economic Action*, New York, NY: Basic Books.

Santos, J. et al., 2016. Does a small cost share reflect a negligible role for energy in economic production? Testing for aggregate production functions including capital, labor, and useful exergy through a cointegration-based method. *Munich Personal Repec Archive*.

Schaller, M., 2010. *Evolution, Culture, and the Human Mind*, New York, NY: Psychology Press.

Schaltegger, C.A. & Torgler, B., 2010. *Work ethic, Protestantism, and Human Capital*,

Schlesinger, A.M., 1986. *The Cycles of American History*, Boston, MA: Houghton Mifflin Company.

Schultz, T., 1964. *Transforming Traditional Agriculture.*, New Haven, CT: Yale University Press.

Schumpeter, J., 1942. *Capitalism, Socialism and Democracy*, London, UK: Harper & Brothers.

Segerstrom, P.S. et al., 1990. A Schumpeterian Model of the Product Life Cycle. *American Economic Review*, 80(5), pp.1077–1091.

Seghetta, M., Østergård, H. & Bastianoni, S., 2014. Energy analysis of using macroalgae from eutrophic waters as a bioethanol feedstock. *Ecological Modelling*, 288(3), pp.25–37.

Sell, B., Murphy, D. & Hall, C.A.S., 2011. Energy Return on Energy Invested for Tight Gas Wells in the Appalachian Basin, United States of America. *Sustainability*, 3(12), pp.1986–

2008.

Shaw, E.S., 1973. *Financial Deepening in Economic Development*, New York, NY: Oxford University Press.

Sieferle, R.P., 1997. Das vorindustrielle Solarenergiesystem. In H. G. Brauch, ed. *Energiepolitik*. Berlin, DE: Springer.

Skinner, B.J., 1976. A Second Iron Age Ahead? *American Scientist*, 64(3), pp.258–269.

Smil, V., 2008. *Energy in Nature and Society: General Energetics of Complex Systems*, Cambridge, MA: MIT Press.

Smil, V., 2010. *Energy Transitions: History, Requirements, Prospects*, Santa Barbara, CA: Praeger Publishers Inc.

Smith, A., 1776. *An Inquiry into the Nature and Causes of the Wealth of Nations*, London, UK: W. Strahan, T.Cadell.

Smulders, S., 1995. Entropy, Environment, and Endogenous Economic Growth. *International Tax and Public Finance*, 2, pp.319–340.

Soam, S. et al., 2015. Life cycle assessment of fuel ethanol from sugarcane molasses in northern and western India and its impact on Indian biofuel programme. *Energy*, 83, pp.307–315.

Soares, R.R., 2005. Mortality Reductions, Educational Attainment, and Fertility Choice. *American Economic Association Mortality Reductions The American Economic Review*, 95(3), pp.580–601.

Soddy, F., 1926. *Wealth, Virtual Wealth and Debt: The Solution of the Economic Paradox*, New York, NY: E.P. Dutton.

Sokoloff, K.L. & Engerman, S.L., 2000. History Lessons: Institutions, Factor Endowments, and Paths of Development in the New World. *Journal of Economic Perspectives*, 14(3), pp.217–232.

Solow, R.M., 1956. A Contribution to the Theory of Economic Growth. *Quaterly Journal of Economics*, 70(1), pp.65–94.

Solow, R.M., 1970. *Growth Theory: An Exposition*, Oxford, UK: Clarendon Press.

Solow, R.M., 1973. Is the End of the World at Hand? *Challenge*, 16(1), pp.39–50.

Solow, R.M., 1957. Technical Change and the Aggregate Production Function. *Review of Economics and Statistics*, 39(3), pp.312–320.

Solow, R.M., 1974. The Economics of Resources or the Resources of Economics. *American Economic Review*, 64(2), pp.1–14.

Spencer, H., 1890. *First Principles*, New York, NY: D. Appleton and Company.

Steen, B. & Borg, G., 2002. An estimation of the cost of sustainable production of metal concentrates from the earth’s crust. *Ecological Economics*, 42(3), pp.401–413.

Stern, D.I., 2004. Economic Growth and Energy. In C. J. Cleveland & R. U. Ayres, eds. *Encyclopedia of Energy, Vol. 2*. Boston, MA: Elsevier Academic Press.

Stiglitz, J., 1974. Growth with Exhaustible Natural Resources: Efficient and Optimal. *Review of Economic Studies*, 41, pp.123–137.

Strauss, W. & Howe, N., 1997. *The Fourth Turning: An American Prophecy*, New York, NY: Broadway Books.

Stringer, C., 2003. Human Evolution: Out of Ethiopia. *Nature*, 423(6941), pp.692–693.

Swan, T.W., 1956. Economic Growth and Capital Accumulation. *Economic Record*, 32(2), pp.334–361.

Swenson, R., 1989. Emergent attractors and the law of maximum entropy production: Foundations to a theory of general evolution. *Systems Research*, 6(3), pp.187–197.

Swenson, R., 2010. Selection Is Entailed by Self-Organization and Natural Selection Is a Special Case. *Biological Theory*, 5(2), pp.167–181.

Swenson, R., 2009. The fourth Law of thermodynamics or the Law of Maximum Entropy Production (LMEP). *Chemistry*, 18(5), pp.333–339.

Tabellini, G., 2010. Culture and Institutions: Economic Develomment in the Regions of Europe. *Journal of the European Economic Association*, 8(4), pp.677–716.

Tahvonen, O. & Salo, S., 2001. Economic growth and transitions between renewable and nonrenewable energy resources. *European Economic Review*, 45(8), pp.1379–1398.

Tainter, J.A., 1988. *The Collapse of Complex Societies*, New York, NY: Cambridge University Press.

Tawney, R., 1926. *Religion and the Rise of Capitalism*, New York, NY: Harcourt Brace and Co.

Tharumarajah, A. & Koltun, P., 2011. Cradle-to-Gate Assessment of Environmental Impact of Rare Earth Metals. In *7th Australian Conference on Life Cycle Assessment*.

The Maddison Project, 2013. GDP per Capita 1-2010 AD. Available at: http://www.ggdc.net/maddison/maddison-project/home.htm.

The Shift Project, 2015. Historical Energy Production Statistics. Available at: http://www.tsp-data-portal.org/Energy-Production-Statistics#tspQvChart.

Thoenig, M. & Verdier, T., 2003. A Theory of Defensive Skill-Biased Innovation and Globalization. *American Economic Review*, 93(3), pp.709–728.

Thompson, W.S., 1938. The Effect of Housing upon Population Growth. *Milbank Memorial Fund Quarterly*, 16(4), pp.359–368.

Thomson, W., 1911. *Mathematical and Physical papers, Vol.V: Thermodynamics, cosmical and geological physics, molecular and crystalline theory, electrodynamics*, Cambridge, UK: Cambridge University Press.

Todd, E., 1983. *The Explanation of Ideology: Family Structure and Social Systems*, New York, NY: Blackwell Publishers.

Toynbee, A., 1934. *A Study of History, Vol III: The Growths of Civilizations*, London, UK: Oxford University Press.

Trefler, D., 1993. International Factor Price Differences: Leontief Was Right! *Journal of Political Economy*, 101(6), pp.961–987.

Tryon, F.G., 1927. An Index of Consumption of Fuels and Water Power. *Journal of the American Statistical Association*, 22(159), pp.271–282.

Tsur, Y. & Zemel, A., 2005. Scarcity, growth and R&amp;D. *Journal of Environmental Economics and Management*, 49(3), pp.484–499.

Turchin, P. & Nefedov, S.A., 2009. *Secular Cycles*, Princeton, NJ: Princeton University Press.

Ullah, I.I.T., Kuijt, I. & Freeman, J., 2015. Toward a Theory of Punctuated Subsistence Change. *Proceedings of the National Academy of Sciences*, 112(31), pp.9579–9584.

United Nations, 2015. Department of Economic and Social Affairs, Population Division, World Population Prospects, the 2015 Revision. Available at: http://esa.un.org/unpd/wpp/Download/Standard/Population/.

United Nations, 1999. *The World at Six Billion. Part 1 - Introduction and Table 1-4*, Available at: http://www.un.org/esa/population/publications/sixbillion/sixbilpart1.pdf.

US Bureau of Labor Statistics, 2016. Labor Force Statistics from the Current Population Survey. Available at: http://data.bls.gov/timeseries/LNU04000000?years_option=all_years&periods_option=specific_periods&periods=Annual+Data.

US Bureau of the Census, 1975a. *Historical Statistics of the United States, Colonial Times to 1970. Bicentennial Edition, Part 1*, Washington, DC. Available at: https://fraser.stlouisfed.org/docs/publications/histstatus/hstat1970_cen_1975_v1.pdf.

US Bureau of the Census, 1975b. *Historical Statistics of the United States, Colonial Times to 1970. Bicentennial Edition, Part 2*, Washington, DC. Available at: https://fraser.stlouisfed.org/docs/publications/histstatus/hstat1970_cen_1975_v2.pdf.

US Department of Energy, 2016. Alternative Fuels Data Center, Fuel Prices. Available at:

http://www.afdc.energy.gov/fuels/prices.html.

US Energy Information Administration (EIA), 2013. *Annual Coal Report 2012*, Available at: http://www.eia.gov/coal/annual/archive/05842012.pdf.

US Energy Information Administration (EIA), 2012. *Annual Energy Review 2011*, Available at: https://www.eia.gov/totalenergy/data/annual/pdf/aer.pdf.

US Energy Information Administration (EIA), 2014. International Energy Statistics database. Available at: http://www.eia.gov/cfapps/ipdbproject/IE.

US Energy Information Administration (EIA), 2016. *Monthly Energy Review (June 2016)*, Available at: http://www.eia.gov/totalenergy/data/monthly/pdf/mer.pdf.

USGS, 2012. Commodity Statistics and Information. Available at: http://minerals.usgs.gov/minerals/pubs/commodity/.

Uzawa, H., 1961. Neutral Inventions and the Stability of Growth Equilibrium. *Review of Economic Studies*, 28(2), pp.117–124.

Valero, A. & Botero, E., 2002. Exergetic Evaluation of Natural Mineral Capital (2): Application of the Methodology to Current World Reserves. In *15th International Conference on Efficiency, Costs, Optimization, Simulation and Environmental Impact of Energy Systems*. Berlin, DE, pp. 62–68.

Ventura, J., 1997. Growth and Interdependence. *Quarterly Journal of Economics*, 112(1), pp.57–84.

Vernon, R., 1966. International Investment and International Trade in Product-Cycle. *Quarterly Journal of Economics*, 80(2), pp.190–207.

Vidal, O., Goffé, B. & Arndt, N., 2013. Metals for a Low-Carbon Society. *Nature Geoscience*, 6(11), pp.894–896.

van der Voet, E. et al., 2013. *Environmental Risks and Challenges of Anthropogenic Metals Flows and Cycles*, United Nations Environment Programme.

Voth, H.-J., 2008. Clark's intellectual Sudoku. *European Review of Economic History*, 12(2), pp.149–155.

de Vries, J., 1994. The Industrial Revolution and the Industrious Revolution. *Journal of Economic History*, 54(2), pp.249–270.

Vries, P., 2015. *State, Economy and the Great Divergence: Great Britain and China, 1680s-1850s*, New York, NY: Bloomsbury Academics.

Vries, P., 2010. The California School and Beyond: How to Study the Great Divergence? *History Compass*, 8(7), pp.730–751.

Wade, N., 2014. *A Troublesome Inheritance: Genes, Race, and Human History*, New York, NY: The Penguin Press.

Walras, L., 1874. *Éléments d'Économie Politique Pure ou Théorie de la Richesse Sociale*, Lausanne, SW: Corbaz.

Warr, B. et al., 2010. Energy use and economic development: A comparative analysis of useful work supply in Austria, Japan, the United Kingdom and the US during 100 years of economic growth. *Ecological Economics*, 69(10), pp.1904–1917.

Warr, B. & Ayres, R.U., 2012. Useful work and information as drivers of economic growth. *Ecological Economics*, 73, pp.93–102.

Warren, G. & Pearson, F.A., 1933. *Prices*, New York, NY: John Wiley and Sons.

Weber, M., 1930. *The Protestant Ethic and the Spirit of Capitalism*, London, UK: Allen & Unwin.

Weil, D.N., 2013. *Economic Growth* 3rd ed., Harlow, UK: Pearson Education.

Weißbach, D. et al., 2013. Energy intensities, EROIs (energy returned on invested), and energy payback times of electricity generating power plants. *Energy*, 52, pp.210–221.

van der Werf, E., 2008. Production functions for climate policy modeling: An empirical analysis. *Energy Economics*, 30(6), pp.2964–2979.

Weyant, J.P., 2011. Accelerating the development and diffusion of new energy technologies: Beyond the “valley of death.” *Energy Economics*, 33(4), pp.674–682.

White, L., 1943. Energy and the Evolution of Culture. *American Anthropologist*, 45(3), pp.335–356.

Williamson, O.E., 2000. The New Institutional Economics: Taking Stock, Looking Ahead. *Journal of Economic Literature*, 38(3), pp.595–613.

Wolpoff, M.H., Hawks, J. & Caspari, R., 2000. Multiregional, Not Multiple Origins. *American Journal of Physical Anthropology*, 112(1), pp.129–136.

Wong, R. Bin, 1997. *China Transformed: Historical Change and the Limits of European Experience*, Ithaca, NY: Cornell University Press.

World Bank, 2016a. World Bank Indicators, GDP growth (annual %). Available at: http://data.worldbank.org/indicator/NY.GDP.MKTP.KD.ZG.

World Bank, 2016b. World Bank Indicators, gross capital formation (% of GDP). Available at: http://data.worldbank.org/indicator/NE.GDI.TOTL.ZS.

World Bank, 2016c. World Bank Indicators, gross savings (% of GDP). Available at: http://data.worldbank.org/indicator/NY.GNS.ICTR.ZS.

Wrigley, E.A., 1988. *Continuity, Chance and Change: The Character of the Industrial Revolution in England*, Cambridge, UK: Cambridge University Press.

Wrigley, E.A., 2016. *The Path to Sustained Growth: England’s Transition from an Organic Economy to an Industrial Revolution*, Cambridge, UK: Cambridge University Press.

Young, A., 1991. Learning by Doing and the Dynamic Effects of International Trade. *Quarterly Journal of Economics*, 106(2), pp.369–405.

Zak, P.J. & Knack, S., 2001. Trust and Growth. *Economic Journal*, 111(470), pp.295–321.

Zarrouk, S.J. & Moon, H., 2014. Efficiency of Geothermal Power Plants: A Worldwide Review. *Geothermics*, 51, pp.142–153.

Zeh, H.D., 2007. *The Physical Basis of the Direction of Time* 5th ed., New York, NY: Springer.

Zhang, J. & Nishimura, K., 1993. The old-age security hypothesis revisited. *Journal of Development Economics*, 41(1), pp.191–202.

Zhang, Y. & Colosi, L.M., 2013. Practical ambiguities during calculation of energy ratios and their impacts on life cycle assessment calculations. *Energy Policy*, 57, pp.630–633.

# LIST OF FIGURES

**Chapter 1**

**Chapter 2**

**Chapter 3**

**Chapter 4**

**Chapter 5**

**Chapter 6**

**Chapter 7**

**Conclusion**

**Appendices**

*Appendix D*

*Appendix G*

*Appendix H*

*Appendix I*

# LIST OF TABLES

**Chapter 1**

**Chapter 4**

**Chapter 5**

**Chapter 6**

**Chapter 7**

## Appendices

*Appendix B*

*Appendix D*

*Appendix I*

*Appendix J*

*Appendix K*

# GLOSSARY

***Anergy***: the useless part of energy that cannot deliver work in the physical sense as opposed to exergy. Anergy grows at the expense of exergy in all real processes because irreversibilities generate the production of entropy.

***Availability***: the magnitude of a stock or flow of primary energy available for humankind. In other words it is the total quantity of a given primary energy resource that may be ultimately recovered by human technology. It is more formally defined by the Ultimately Recoverable Resource (URR, generally in EJ) for nonrenewable energy resource, and by the Technical Potential (TP, generally in EJ/year)

***Accessibility***: the degree of difficulty associated with the extraction of given primary energy type from the environment. In a biophysical perspective it is represented by the energy-return-on-investment (EROI) of energy resources. Mainstream economists prefer to use the relative costs of energy technologies as a measure of this same variable.

***Attainability***: the maximum possible level that aggregate technology of the economy will eventually reach in the future. It will be recalled that in a biophysical perspective the macroeconomic technological level is defined as the efficiency of primary-to-useful exergy conversion.

***Cost share theorem***: the principle according to which output elasticities of factors of production are equal to their respective shares in the total macroeconomic output. Generally, labor wages and rents from capital respectively represent 70% and 30% of total national income in industrialized countries. As a consequence, their output elasticities (i.e. marginal productivities) are set to 0.7 and 0.3 respectively. This neoclassical theorem is in fact completely misleading and proved wrong when one adopts a biophysical perspective that takes into account technological constraints.

***Demographic transition***: the transition from high birth and death rates to lower birth and death rates as a country develops from a pre-industrial to an industrialized economic system.

***Dissipative structure***: a dissipative system that has a dynamical regime that is in some sense in a reproducible steady state. This reproducible steady state may be reached by natural evolution of the system, by artifice, or by a combination of these two. A dissipative structure is characterized by the spontaneous appearance of symmetry breaking (anisotropy) and the formation of complex, sometimes chaotic, structures where interacting particles exhibit long range correlations. The term dissipative structure was coined by Russian-Belgian physical chemist Ilya Prigogine, who was awarded the Nobel Prize in Chemistry in 1977 for his pioneering work on these structures.

***Dissipative system***: a thermodynamically open system which is operating out of, and often far from, thermodynamic equilibrium in an environment with which it exchanges energy and matter.

***Economic growth***: the increase in the inflation-adjusted market value of the goods and services produced by an economy over time. It is conventionally measured as the percent rate of increase in real gross domestic product, or real GDP, usually in per capita terms.

***Economic system***: a system of production, resource allocation, exchange, and distribution of goods and services in a society or a given geographic area. It includes the combination of the various institutions, agencies, entities, decision-making processes, and patterns of consumption that comprise the economic structure of a given community. As such, an economic system is a type of social system. All economic systems have three basic questions to ask: what to produce, how to produce and in what quantities, and who receives the output of production. One part of the economic system, the energy system, produces energy, whereas the rest of the economy dissipates energy to maintain its functioning. In this sense, apart from the energy system, the economy is a dissipative structure.

***Elasticity of substitution***: measures the percentage change in the ratio of two inputs used in response to a percentage change in their prices. It measures the curvature of an isoquant and thus, the substitutability between inputs (or goods), i.e. how easy it is to substitute one input (or good) for the other.

***Endogenous vs. exogenous***: refer to the nature of changes in economic models. Endogenous changes are those that originate from within the model functioning, whereas exogenous changes come from outside the model and are therefore unexplained by the model. One will also talk about endogenous variables of the model and exogenous parameters (i.e. constants).

***Energy***: measures of the ability of a body or system to do work or produce a change, expressed usually in joules (J) or kilowatt hours (kWh). No activity is possible without energy and its total amount in the universe is fixed. In other words, it cannot be created or destroyed but can only be changed from one type to another.

***Energy expenditures***: (or energy cost) is the quantity of economic output that must be allocated to obtaining energy. It is usually expressed as a fraction of the gross domestic product (GDP).

***Energy intensity***: is a measure of the energy efficiency of a nation’s economy. It is calculated as units of energy (e.g. MJ) per unit of GDP (e.g. $1990), so that generally it is expressed in MJ/$1990.

***Energy resource***: is any natural capital of use for the production of energy. Energy resources are divided into nonrenewable stocks of fossil fuels (coal, oil, and gas) and nuclear fuels (uranium, thorium, etc.), and renewable resources flows (biomass, solar, wind, hydro, geothermal, wave, and tidal). The estimates for the amount of energy in these resources is usually given in exajoules (EJ = $10^{18}$ J), it represents their availability .

***Energy system***: is the part of the economic system that extracts primary energy from the environment and refined it in the forms of final carriers dissipated by the rest of the economy system.

***Energy-Economy System***: is the association of the energy-producing energy system with the energy-dissipating economic system.

***Entropy***: is a measure of the number of microscopic configurations Ω that correspond to a thermodynamic system in a state specified by certain macroscopic variables. Thermodynamic entropy is a state function (meaning that it does not depend on the path by which the system arrived at its present state) that represent the level of disorder of the system. The higher the entropy of a system, the higher its entropy.

***Energy-return-on-investment (EROI)***: is the ratio of the amount of usable energy delivered from a particular energy resource to the amount of energy used to obtain energy from that resource. A potential energy resource is interesting only if it delivers more energy than what has been invested to exploit it, meaning that its EROI must be superior to unity.

***Exergy***: is the maximum useful work possible during a process that brings the system into equilibrium with a heat reservoir. When the surroundings are the reservoir, exergy is the potential of a system to cause a change as it achieves equilibrium with its environment. Exergy is the energy that is available to be used. Exergy is always destroyed when a process involves a temperature change and the destroyed exergy is called anergy. The destruction of exergy is proportional to the entropy increase of the system together with its surroundings.

***Exploited resource ratio***: is a measure of how much of a given energy resource is currently being produced, normalized with respect to the total availability of that same resource. In the case of nonrenewable energy, the exploited resource ratio is defined as the cumulated production divided by the ultimately recoverable resource (URR). In the case of renewable energy, the exploited resource ratio is defined as the annual production divided by the technical potential (TP). In both cases the exploited resource ratio equals 0 when the energy resource under study is virgin, and it equals 1 when it is fully depleted.

***Factors of production***: are what is used in the production process to produce output—that is, finished goods and services. In mainstream economics, there are two basic resources or factors of production: labor and capital. Generally, the changing quality of these factors must be taken into account and their combined growth cannot match output growth so that a variable representing technological change must be included in production function. The major flaw of mainstream economics is to have disregard exergy as the major factor of production of the economy.

***General Purpose Technology***: can affect an entire economy (usually at a national or global level) with the potential to drastically alter societies through its impact on pre-existing economic and social structures. Examples include the steam engine, railroad, interchangeable parts, electricity, electronics, material handling, mechanization, control theory (automation), the automobile, the computer, the Internet.

***Great Divergence***: refers to the process by which the Western world (i.e. Western Europe and the Western Offshoots, which the parts of the New World where European natives became the dominant populations) overcame pre-modern growth constraints and emerged during the nineteenth century as the most powerful and wealthy regions of all time, eclipsing Qing China, Mughal India, Tokugawa Japan, and the Ottoman Empire. The term is sometimes misleadingly attributed to Samuel Huntington (in particular in Wikipedia), but there is no trace of this very name in his work.

***Gross domestic product (GDP)***: is a monetary measure of the market value of all final goods and services produced in a period (quarterly or yearly). Nominal GDP estimates are commonly used to determine the economic performance of a whole country or region, and to make international comparisons. Nominal GDP, however, does not reflect differences in the cost of living and the inflation rates of the countries; therefore using a GDP PPP per capita basis is arguably more useful when comparing differences in living standards between nations.

***Human capital***: refers, in the Becker-Mincer view, to the stock of knowledge, habits, social and personality attributes, including creativity, embodied in the ability to perform labor so as to produce economic value. The Nelson-Phelps-Schultz approach is different and emphasized that the major role of human capital is not to increase productivity in existing tasks, but to enable workers to cope with change, disruptions, and the implementation of new technologies.

***Joule***: a derived unit of energy in the International System of Units. It is equal to the energy transferred (or work done) to an object when a force of one newton acts on that object in the direction of its motion through a distance of one meter (1 newton meter or N·m). It is also the energy dissipated as heat when an electric current of one ampere passes through a resistance of one ohm for one second.

***Labor***: a factor of production representing the amount of work done by human beings. It is generally measured in hours of work per year.

***Malthusian Epoch***: constitutes the state in which mankind relied for most of its history, from 150,000 BCE to 1825 CE. In the Malthusian Epoch, or Malthusian Trap, income was largely stagnant because technological advances and discoveries only resulted in more people, rather than improvements in the standard of living.

***Modern Growth Era***: is a state of sustained high economic growth attained around 1950 at the global scale but several decades before by leading countries such as the USA, the UK, France, and Germany. In such a state, population growth no longer offsets the rise in aggregate income that is enabled by ever increasingly efficient use of accumulating production factors.

***Net energy***: refers to the difference between the energy expended to harvest an energy source and the amount of energy gained from that harvest. A potential energy resource is interesting only if it delivers more energy than what has been invested to exploit it, so that its net energy gain must be superior to zero.

***Post-Malthusian Regime***: is short-term state in which countries escaping the Malthusian Trap rely for several decades. At some point in the early process of industrialization, the level of capital accumulation and technological progress allowed income per capita to rise by counteracting the effect that higher population growth had on diluting the economic product in the Malthusian Epoch. In the Post-Malthusian Regime, the fertility rate, birth rate, and death rate all declined compared to the Malthusian Epoch, whereas life expectancy, literacy rate, industrialization and urbanization levels increased.

***Production function***: relates physical output of a production process to physical inputs or factors of production. The primary purpose of the production function is to address allocative efficiency in the use of factor inputs in production and the resulting distribution of income to those factors. In macroeconomics, aggregate production functions are estimated to create a framework in which to distinguish how much of economic growth to attribute to changes in factor allocation (i.e. the accumulation of capital and labor) and how much to attribute to advancing technology.

***Physical capital***: a factor of production (or input into the process of production), consisting of machinery, buildings, computers, and the like. Capital goods, real capital, or capital assets are already-produced durable goods or any non-financial asset that is used in production of goods or services.

***Qualitative depletion***: is the degradation of the average quality of a (non)renewable stock. While quantitative depletion represents the consumption of a resource faster than it can be replenished, qualitative depletion represents the fact that the average grade of this resource also decreases with the resource extraction.

***Technical Potential (TP)***: is the total renewable resource recoverable by a specified technology.

***Technological change***: is a change in the set of feasible production possibilities in mainstream economics. Technological change is Harrod neutral if the technology is labor-augmenting (i.e. increases labor productivity); it is Solow neutral if the technology is capital-augmenting (i.e. increases capital productivity); it is Hicks neutral if the technology is both labor and capital-augmenting (i.e. increases labor and capital productivity in the same proportion). In such a mainstream view, technological change is loosely defined and correspond to the catch-all aggregation of primary-to-final and final-to-useful energy conversion efficiency (if energy is considered as an input factor), the division and organization of labor, the broader organization and efficiency of markets, the skill improvements of laborers, the contribution of information and communication technologies, but also the beneficial effects of inclusive institutions (which, for example, protect private property rights and consequently incentivize innovation and R&D). With a biophysical approach of the economic system, technological change is more formally defined as gains in the aggregate efficiency of primary-to-useful exergy conversion.

***Ultimately recoverable resource (URR)***: is the part of a nonrenewable resource stock, including discovered, undiscovered, produced and unproduced amounts, that is producible under a specified set of costs and technologies.

***Unified Growth Theory (UGT)***: is currently being developed by Oded Galor and his co-authors to characterize in a single dynamical system an initial stable Malthusian equilibrium which due to the evolution of latent state variables, ultimately vanishes endogenously, causing a transitional growth take-off before the system gradually converges to a modern growth steady-state equilibrium. Unified growth theory suggests that the transition from stagnation to growth has been an inevitable by-product of the process of development. It argues that the inherent Malthusian interaction between the rate of technological progress and the size and composition of the population accelerated the pace of technological progress and ultimately raised the importance of education in coping with the rapidly changing technological environment. The rise in industrial demand for education brought about significant reductions in fertility rates. It enabled economies to divert a larger share of the gains from factor accumulation and technological progress to the enhancement of human capital formation and income per capita, paving the way for the emergence of sustained economic growth. The theory further explores the dynamic interaction between human evolution and the process of economic development and advances the hypothesis that the forces of natural selection played a significant role in the evolution of the world economy from stagnation to growth. The Malthusian pressures have acted as the key determinant of population size and conceivably, via natural selection, have shaped the composition of the population as well. Lineages of individuals whose traits were complementary to the economic environment generated higher levels of income, and thus a larger number of surviving offspring, and the gradual increase in the representation of their traits in the population contributed to the process of development and the take-off from stagnation to growth. UGT is probably the most promising analytical framework to study and understand the phenomenon of Great Divergence but it disregard the biophysical reality of the economy system and currently violates the first and second laws of thermodynamics. UGT will necessarily have to comply with these laws to successfully describe the evolution of mankind.