D.-Ing. Arnulf Grübler
Wiedner Gürtel 12/III/14 a
1040 Wien

Growth to Limits

Long Waves and the Dynamics of Technology

Nebojsa Nakicenovic

Vienna • 1984

D.-Ing. Arnulf Grübler
Wiedner Gürtel [illegible]/III/14 a
1040 Wien

Growth to Limits

Long Waves and the Dynamics of Technology

ABSTRACT

Economic development and technological change are analyzed as a process of substituting old forms of satisfying human needs by new ones. Logistic substitution analysis is used to show the dynamics and regularity of technological changes. These technological changes are related to the long waves or swings in economic development and many other indicators ranging from energy intensity to price fluctuations and innovation pulses. A phenomenological approach is adopted to evaluate the evidence for invariance and logical order in the sequence of technological changes and long wave fluctuations. The analysis is based on the longest historical time series for the United Kingdom, the United States and the world that can be reconstructed from the sparse records.

Growth to Limits

Long Waves and the Dynamics of Technology

Nebojsa Nakicenovic

Vienna • 1984

DRAFT
NOT FOR QUOTATION
WITHOUT PERMISSION
OF THE AUTHOR

Growth to Limits

Long Waves and the Dynamics of Technology

ABSTRACT

Economic development and technological change are analyzed as a process of substituting old forms of satisfying human needs by new ones. Logistic substitution analysis is used to show the dynamics and regularity of technological changes. These technological changes are related to the long waves or swings in economic development and many other indicators ranging from energy intensity to price fluctuations and innovation pulses. A phenomenological approach is adopted to evaluate the evidence for invariance and logical order in the sequence of technological changes and long wave fluctuations. The analysis is based on the longest historical time series for the United Kingdom, the United States and the world that can be reconstructed from the sparse records.

PREFACE

The major objective of this work is to investigate the pattern of economic and technological changes during the last two to three centuries. Most of the examples that could be reconstructed from historical records describe the quantitative changes in the overall economic activity, the energy system, the steel industry, marine transport, patents and innovations in the United Kingdom, the United States and the world. The major conclusion is that the events that characterize profound changes in technology and economic structure are timed over intervals of about fifty years and that the pattern of their recurrence is invariant. This conclusion leaves many questions open for further research, but it indicates a promising direction. Perhaps the most important question concerns the explanation of the fifty-year period that times dynamic changes in technology, long waves in economic development and innovation pulses.

This study evolved as a doctoral dissertation at Universität Wien under the supervision of Prof. G. Bruckmann, Universität Wien, and Prof. P.-J. Jansen, Technische Universität Wien.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

Page

Growth to Limits

Long Waves and the Dynamics of Technology

INTRODUCTION

During the three decades after the Second World War the industrialized countries experienced a unique phase of rapid growth and prosperity. Thus, it is not surprising that the expectations and even research about the future were based on the secular trends of these three decades. It was not until more serious economic, social and environmental concerns caused a more cautious view about the realistically achievable futures that some researchers also adopted a longer historical perspective for the analysis of future developments. The economic stagnation of the 1970s, which was associated with the abrupt crude oil price increases, and the first signs of a more serious and long-term depression in the early 1980s, revived interest in the nature of numerous fluctuations that are inherent in economic growth. These fluctuations are not just the manifestation of more and less growth during various phases of economic development. Especially the long waves in economic development that overlap the shorter-term business cycles appear to be associated with fundamental technological and structural changes throughout the economy and even the whole society.

The first part of this work will analyze the important technological changes that have occurred since the beginning of the industrial age. In particular, the generalized logistic substitution model will be used to show the dynamics of these changes. The second part will try to show that technological substitution and the long waves in economic development are not only related by chance, but that they are both manifestations of the same basic processes. Especially the long term fluctuations in the general price level with sharp rises during stagnation, after long periods of growth and prosperity, appear to be very closely connected with technological substitution processes.

Technological Substitution

Economic development and the advancement of technology can be viewed as a process of substituting old forms of satisfying human needs by new ones. This process is very complex and there exists a certain danger of oversimplification in emphasizing only some of its characteristics that appear to be important. It is certainly true that the substitution processes are not only based on the relative superiority of introduced innovations over the old practice in some technical sense, but that the old is sometimes no longer acceptable for some social or political reasons. Be that as it may, we will advance the view that most of the processes of technological change and economic development in general can be represented as substitutions by a new technology or practice for an established one. Thus, technological substitution plays a fundamental role in the analysis of structural changes that lead to new economic patterns and forms.

In this sense, the technological substitution processes may account for the long swings in economic development by causing major and periodic fluctuations in the historical rate of technological change and, according to this hypothesis, also in the rate of economic growth. The periodicity of the long swings in economic development (i.e., the Kondratieff "long cycle") could be related to the regularity of the technological substitution processes. We will investigate how realistic such a hypothesis is and to what extent it can be supported by empirical evidence.

A typical technological substitution process is preceded by a basic invention that provides the scientific and technical feasibility, and later by a basic innovation in the form of a new economic or social process. If the new technology or product becomes viable, the initial market penetration is followed by a phase of rapid growth with relatively high and assured returns, finally ending in a phase of saturation characterized by relatively low capacity utilization that may culminate in eventual decline. We will discuss each of these different phases of technological change.

We will also review the literature that deals with logistic substitution processes starting with the work of Mansfield, Fisher and Pry, and others, dealing briefly with both the theories that were advanced to explain these processes and their empirical base.

The Generalized Technological Substitution Model

Previous substitution analyses presented a somewhat simplified view of the process of technological change. Although we still maintain that the process of technological change could be characterized as the substitution of old by new technologies or products, more realistically this process is a sequence of such substitutions. Furthermore, the sequential character of various substitution processes does not only mean that the competing technologies or products struggle for "control of the same market", but also that it is possible that more than two competitors could substitute each other simultaneously.

In the case of more than two competitors the usual substitution models are inadequate so that a generalized model must be adopted. Our model will be introduced on the basis of energy, steel and some other examples. In particular, emphasis will be given on the adequacy of this method to "organize" historical data and to project the substitution dynamics into the future. The original projections with the model were performed about six years ago when the most recent data available were for the year 1974. Today data are available up to 1981 and even 1982, so that it is possible to verify the validity of these *ex ante* projections over a period of seven to eight years. This kind of validation of the model will be presented for energy substitution in the United Kingdom, the United States and the world.

The stress of the analysis will be to capture the regularity of the substitution dynamics in these examples. This regularity and stability is captured by the concept of "takeover time", defined as the number of years it takes before a given technology or product can expand its market share from one to fifty percent, provided of course that it turns out to be successful.

Long Swings and Secular Movements

The fact that prices show secular movements of a longer period than the usual business cycles has already been known for some time. This phenomenon received wide attention in the 1920s following the appearance of the work of Kondratieff (also transliterated from Russian as Kondratiev) on the "long cycles" of economic development. Notably, Schumpeter, Kuznets, Mitchell, Burns and others attempted to provide theoretical formulation and more empirical evidence for and against the long waves of price and physical output movements. After the Second World War their research was almost forgotten, probably due to the historically unique period of growth and prosperity enjoyed in the industrialized countries. During the last decade, ever since the onset of a more serious and probably longer lasting slowdown of the economies throughout the world, the notion of long waves or "cycles" has attracted much interest and revival as a topic of research.

Up to now there have been at least four different basic approaches in explaining the long waves phenomenon that vary from "bunching" of basic innovations to "lumpiness" of capital goods and subsequent "echos" of these effects in the form of long waves. Some of these theories account also for "external" causes and in others they are considered to be "internal" to the nature of capitalistic growth itself. We will try to show that both long waves and technological substitution are closely related to each other and also to the entrepreneurial and technological innovations and their acceptance. We are however not in a position to identify any "causality rules" given our phenomenological approach. All we can do is to identify the secular movements and their relationships, and once they are clearly understood a theoretical formulation should be within reach.

Historical Data and Their Secular Trends

The discussion will center around two aspects of historical data that are available in the form of time series - their secular trends and fluctuations. These two "components" are in principle indistinguishable, but statistical methods can make their separation possible. In this kind of analysis it is necessary to first determine the secular trend of the series, since most of the indicators were not stationary during the last two to three hundred years and usually portray strong growth trends in addition to the overlapping fluctuations.

We will show some of the longest time series on prices and physical production and consumption for the United Kingdom, the United States, and the world that we could "splice" together or reconstruct from the sparse literature on economic activities during the last two or three centuries. In some cases, such as the price of wheat, the historical record can be reconstructed extending back more than five centuries. These time series will also be shown in order to illustrate the differences and similarities between the price movements of the industrialized and preindustrialized development phases.

We will confirm Kondratieff's original finding that the long swings can be observed in most of the price series, such as the wholesale price index, interest rates or monetary values of various goods and services, and that it is possible to show similar swings in a very limited number of physical production or consumption figures. Some authors doubt the existence of long swings based on this discrepancy in the empirical evidence. They claim that the long swings are purely a price phenomenon with almost no effects on the "physical" structure of the output. We will challenge this view and attempt to illustrate that some of the physical indices, such as primary energy consumption in the United States, do portray long wave movements although they are not as "cyclical" as similar swings in price indices. The long waves will also be related to the "basic" innovations and their apparent clustering around the slumps of the long waves. Especially the arguments and evidence offered by Mensch for the causality of long waves from the bunching of basic innovations will be discussed in view of our observations.

Technological Substitution and Long Waves

We will basically summarize the empirical evidence shown so far for the regularity of the substitution processes and the alleged periodicity of the long waves. We will show to what extent a "synchronization" can and cannot be observed between these two phenomena. In addition, we will attempt to relate this observation to other "explanations" of the long wave as advanced by Kondratieff and his followers, and also to those which doubt the existence of periodic waves and explain the observed phenomena as unique historical incidents each based on different causes such as the large gold discoveries during the last century.

This analysis will strengthen the hypothesis that the phenomenological evidence for the regularity and periodicity of the long swings in economic development in industrialized countries cannot be explained by any simple cause, because of the complexity of so many different fundamental principles and unique events that remain hidden behind the aggregate level of available historical data. It appears that most of the explanations of long waves and technological change, starting with the clumping of capital accumulation needs or bunching of basic innovations going all the way to the need to replace large infrastructures, have their individual merits and validity so that an overall theory must include many of these principles and not only rely on the mono-causal explanations advanced up to date. Schumpeter is perhaps the most important exception to this practice in that he considered many general and also specific reasons for each of the major swings experienced in the industrialized countries. Today, his work is usually associated with the "innovation hypothesis" and the strategic role of the entrepreneur whose innovations lead to the great waves in economic development.

In that sense this analysis will also only provide another piece of the puzzle and not a solution to the lack of a theoretical base for a more effective explanation of this so fundamental but also complex process. The analysis will show the empirical evidence for and against a relationship between the process of technological change and long swings in the economic development of the industrialized core countries. We will conclude that we need better empirical evidence and understanding of the secular movements and their relationships, in addition to those compiled up to date, before "causality rules" can be formulated that could explain the complex relationships observed today between various indicators. Causality rules are a prerequisite for a consistent explanation (or rather prediction) and a theory.

1 DYNAMICS OF ENERGY SUBSTITUTION

1.1 Technological Change

Substituting an old way by a new one to satisfy a given need has been the subject of a large number of studies. Most of these studies analyzed the substitution of products or technological processes. One general finding is that almost all substitutions of an old product or technology by a new one, expressed as fractional market shares, follow characteristic S-shaped curves with very rapid substitution rates at the beginning and decreasing rates toward the end of the market takeover. Typical examples include the substitution of soaps by detergents, corn by hybrid corn seeds, black-and-white by color TV sets, and various manufacturing processes for making turpentine, paints, steel and so on.

Most of the studies of technological substitution are based on the use of the logistic curve to represent the substitution process. The logistic function, however, is not the only S-shaped function, but it is perhaps the most suitable one for empirical analysis of growth and substitution processes because both the shape and the form of the function correspond to the theoretical explanation of the S-shaped growth processes. The function is symmetrical around its point of inflection and the *relative* rate of increase (actual growth-rate over the achieved growth) declines linearly with increasing level of growth. In other words, the growth-rate is proportional both to the achieved growth-level and the possible growth left to the asymptotic level; the growth-rate increases up to the inflection point and decreases afterwards. In addition, the parameters of the function represent meaningful attributes of the growth process - the growth-rate, the asymptotic growth-level and the location of the function in time. Another S-shaped function, the Gompertz curve, has also been frequently used, particularly to describe population, plant, and animal growth (see, e.g. Richards, 1959). The cumulative normal distribution function has been also applied to describe technological substitution dynamics (see, e.g. Stapleton, 1976). The mathematics of such an approach is more complex than in the case of the logistic or Gomperz curve because the normal distribution function cannot be integrated analytically.

The widespread empirical applications of the logistic function as a means to describe growth phenomena originated in studies of human population, biology and chemistry. The first reference to the logistic function can be found in Verhulst (1838, 1845, 1847). Pearl (1924, 1925) rediscovered the function and used it extensively to describe the growth of population, both human and biological. Since then, numerous studies have been conducted, only to confirm the logistic property of most growth processes. Robertson (1923) was the first to use the function to describe the growth process of a single individual. Later, the function was advanced for use in bio-assay (see, e.g. Emmens, 1941; Wilson and Worcester, 1942; and Bergson, 1944), in studies of the growth of bacterial cultures in a feeding solution, for autocatalytic chemical reactions, and so forth.

Griliches (1957), in his study of the diffusion of the hybrid corn seed in the United States, was one of the first to use the S-shaped curve to describe technological substitution. He showed that hybrid corn replaced traditional corn seed in different States in a very similar way; the S-shaped substitution only being displaced in time by a few years and lasting for longer or shorter periods in different States.

Following the work of Griliches, Mansfield (1961) developed a model to explain the rate at which firms follow an innovator. He hypothesized that the adoption of an innovation is related positively to the profitability of employing the innovation and negatively to the expected investments associated with its introduction. Mansfield substantiated the theoretical implications of his model by analyzing the diffusion of twelve industrial innovations in four major industries.

Mansfield's findings were further extended by Fisher and Pry (1970), who considered only fractional shares of a market controlled by two competing technologies. They postulated, on the basis of their analysis of many substitution processes, that the rate of fractional adoption of a new technology is proportional to both the fraction of the market penetrated by the new technology and the fraction of the market held by the still-used old technology. They also assumed that this substitution process proceeds to complete market takeover by the new technology once it has progressed as far as a few percent of the market. These two basic assumptions describe a growth process that can be represented by a two-parameter logistic function:

$$\frac{f(t)}{1-f(t)} = \exp(\alpha t + \beta)$$

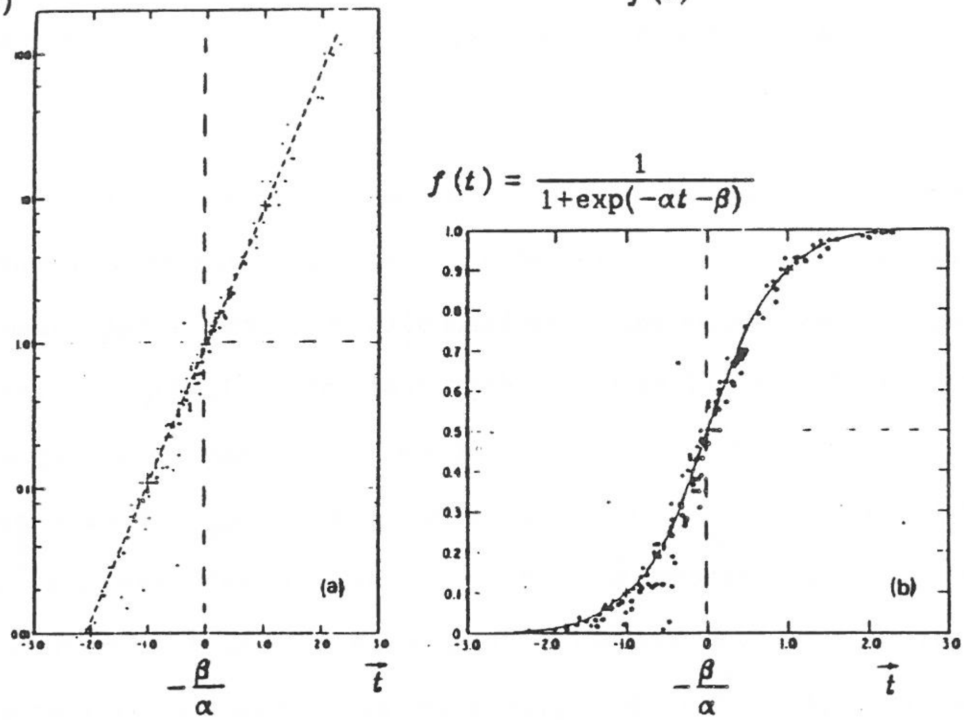
where t is the independent variable usually representing some unit of time, α and β are constants, f is the fractional market share of the new competitor, and $1-f$ that of the old one. The parameters α and β are sufficient to describe the whole substitution process. They cannot be directly observed; they can, however, be estimated from historical data.

The characteristics of the logistic function describing fractional substitution are illustrated in Figure 1.1 which shows the seventeen substitution cases studied by Fisher and Pry and the logistic fit of these data as a smooth line. The plot on the right-hand side of the figure shows the S-shaped form of the function and the scatter of the observed data around the trend line. The plot on the left-hand side

$$\frac{f(t)}{1-f(t)} = \exp(\alpha t + \beta)$$

Differential Form: $$\frac{\dot{f}(t)}{f(t)} = \alpha[1-f(t)],\ 0 < f(t) < 1$$

$$f(t) = \frac{1}{1+\exp(-\alpha t - \beta)}$$



Fit of logistic function to substitution data for seventeen cases vs. normalized units of time. Source: Adopted from Fisher and Pry (1971).

Figure 1.1 Growth to Limits: Seventeen Substitution Cases.

shows another convenient property of the logistic function for empirical analysis. When a sample of data points is plotted as the transformation $f/(1-f)$ on the logarithmic scale, the scatter of observed fractional shares (f) has a linear secular trend, provided that the data points can be described by the logistic function. In the plot on the left-hand side this is the case, the straight line being the logistic fit of the data. On both curves the time has been normalized at the inflection point ($t=-\alpha/\beta$) of the curve, where the fractional share equals one-half and where the slope of the curve is the steepest.

Two sets of examples are shown here in Figure 1.2 and Figure 1.3 from the original papers of Fisher and Pry (Fisher and Pry, 1970; Pry, 1973). Figure 1.2

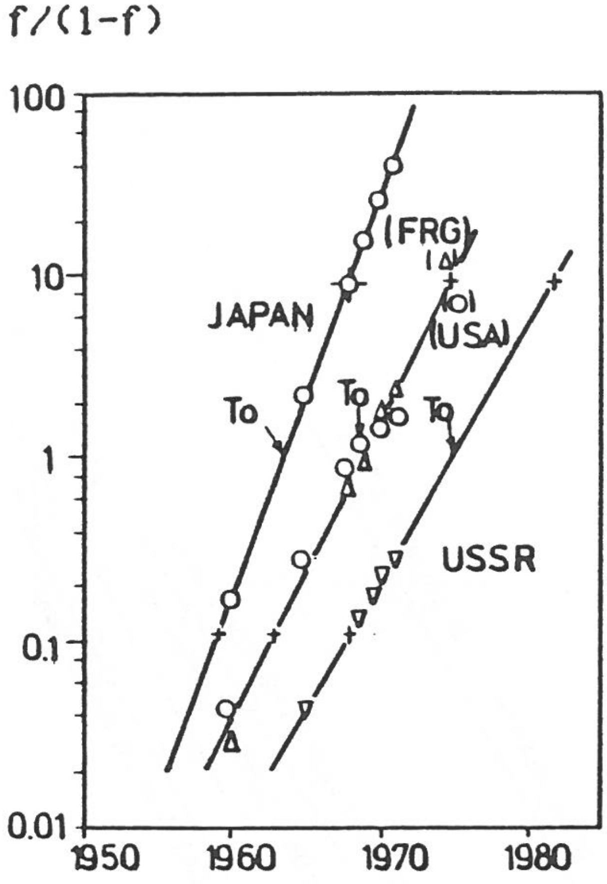


Figure 1.2 Substitution in Steel Production (Pry, 1973).

shows the substitution of basic oxygen furnace for open-hearth and Bessemer steel production expressed by fractional market share (f) for Japan, the Federal Republic of Germany, the United States and the Union of Soviet Socialist Republics. The triangles and the circles on the middle line represent the Federal Republic of Germany and the United States, respectively. Figure 1.3 shows technological substitution, also expressed by fractional market shares (f), in the production of steel, turpentine, and paint. These two sets of examples illustrate that the logistic functions appear to give an excellent description of substitution, not only for very different products and technologies, but also for different types of economies.

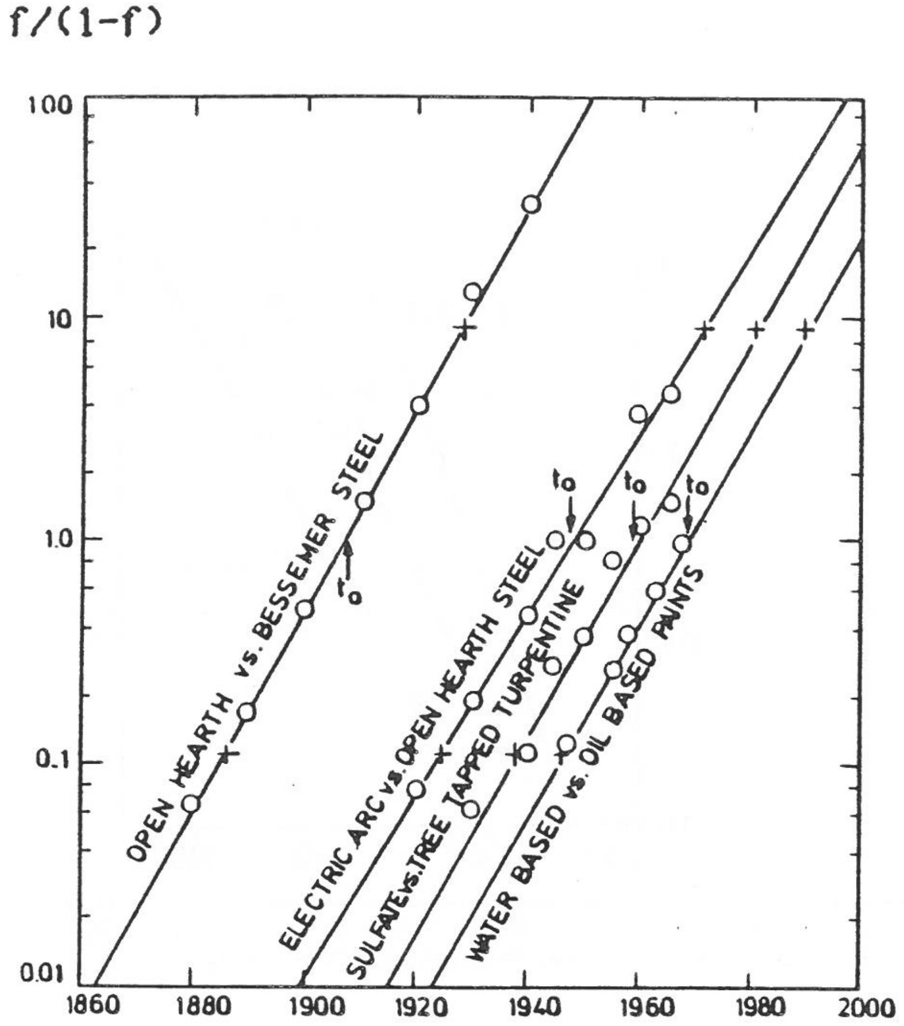


Figure 1.3 Steel, Turpentine and Paint Substitution (Fisher and Pry, 1971).

These examples from the work of Fisher and Pry show the remarkable regularity of technological substitution processes. An obvious next step is to investigate whether the regularity of the substitution dynamics can be applied to more broader technological, economic and social changes. Substitution of various forms of energy offers a good example in this context because energy represents one of the most important inputs for overall economic activity, and because it is used universally in everyday life, it also indicates changes in the acceptance of new energy systems. Energy substitution is representative of large changes at the whole infrastructure level, and thus goes beyond single technological improvements. A new energy source requires changes in accepted practice that range from exploiting new resources of energy, converting them to more convenient energy forms such as electricity and fuels, and transporting and distributing them to the point of end-use. At the end-use level new devices and practices are also needed to convert the delivered energy into the services that are required in various sectors of the economy such as motive power, lighting, space conditioning and so on. Substitution of one form of energy for another implies structural changes that go beyond the changes needed within the energy system because it affects and is in turn affected by the overall economic activities and the whole social structure.

1.2 Global Energy Consumption

During millenia, dating back to the dawn of human civilization, the main sources of energy were fuelwood, wastes (dung and agricultural wastes), and human and animal muscle power. Major exceptions constitute the use of sailing vessels, river flotation, hydro and wind mills. In a wider sense, solar energy was also used, although not directly, but in transformed forms such as dried food and other artifacts. When compared with contemporary energy consumption all of these traditional energy forms were used at low absolute levels of exploitation and low densities of generation and end-use with hardly any need for transportation or transformation. This practice did not only prevail over thousands of years, but it was also basically similar even in different cultural settings. Major variations in energy use were spatially governed. Different geographic and climatic environments imposed different energy use patterns.

With the emergence of manufacturing, industrial production, and the recent rapid social and technological changes over the past few centuries, the energy use patterns have also been altering and improving, energy generation and use densities have increased, and total energy consumption has been growing. Figure 1.4 shows that during the previous 120 years, global primary energy consumption (including the use of fuelwood) increased exponentially at an average growth rate of 2.3 percent per year. Yet during this period energy consumption did not draw equally from all sources, nor did the use of all energy sources increase equally.

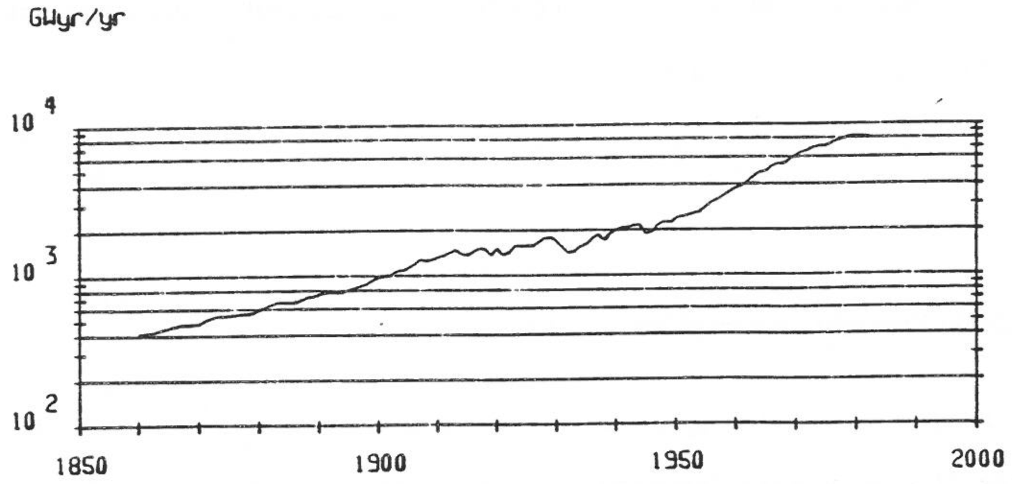


Figure 1.4 Primary Energy Consumption, World.

The next two figures show global primary energy consumption since 1860, according to the five major primary energy sources - fuelwood, coal, crude oil, natural gas and nuclear energy. Figure 1.5 gives consumption according to these five energy sources on a linear scale; Figure 1.6 gives consumption on a logarithmic scale. It is evident from these two plots that the initial growth in the

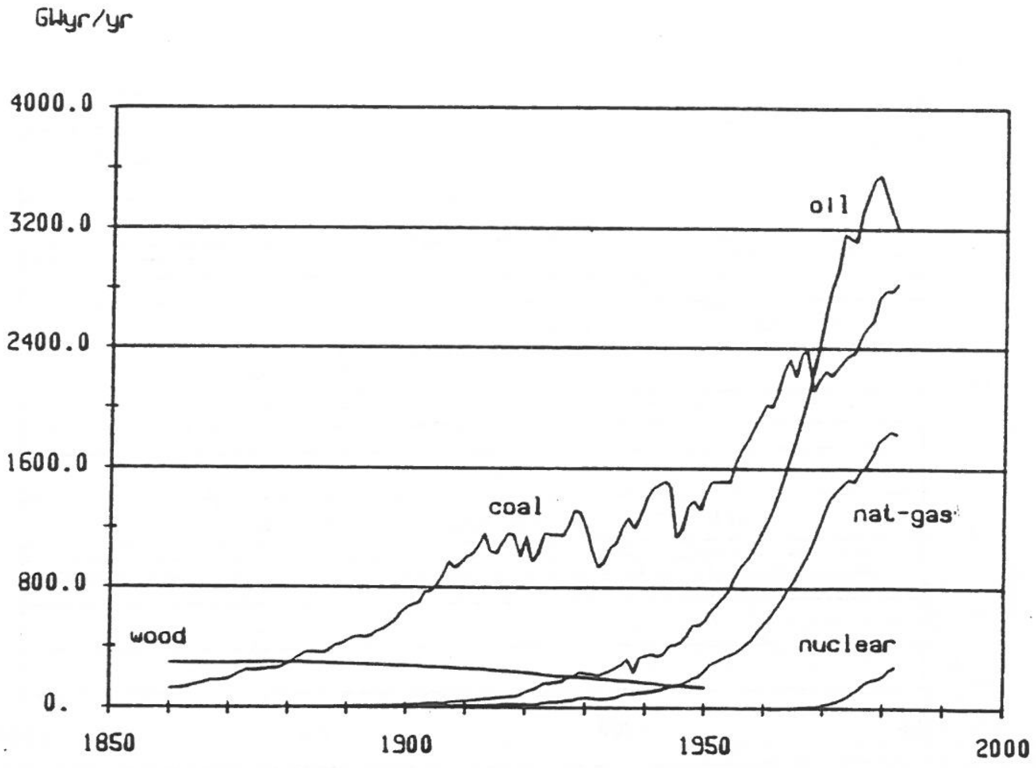


Figure 1.5 Major Primary Energy Sources, World (linear scale).

use of every energy source except fuelwood is exponential (i.e., the secular trend is linear on the logarithmic plot, see Figure 1.6), but that many features, apparently related to economic and political events, influenced energy consumption. During the time when coal held the largest share of the market its consumption was subject to great fluctuations that coincide with the two world wars and the intervening period of worldwide economic depression. The consumption of fuelwood, once the most important source of energy, has decreased since the beginning of the century, although its use is still widespread, especially in the developing parts of the world. In 1965, oil surpassed coal, and natural gas will probably close up in a few years. In fact, oil and natural gas curves have the same shape and almost identical growth rates, they are just shifted in time by about ten to fifteen years. Nuclear energy is still in its early

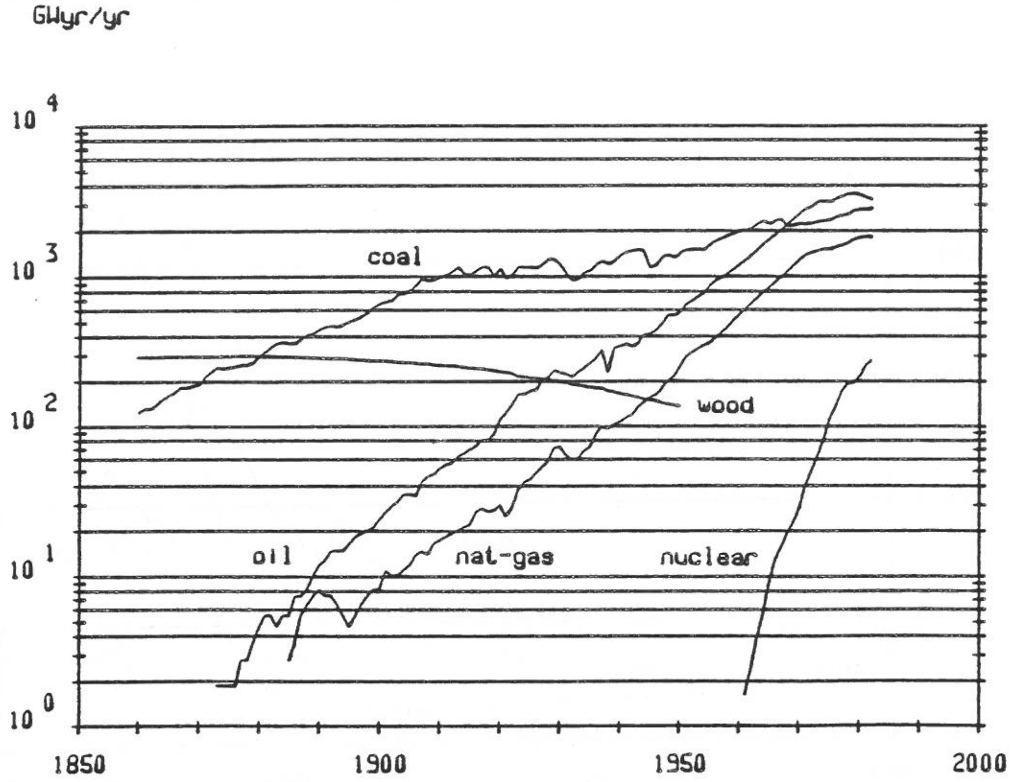


Figure 1.6 Major Primary Energy Sources, World (logarithmic scale).

phase of development, therefore the steep growth prevailing over the last decade may not be indicative of its future role.

Other forms of primary energy have only marginal shares of the total energy consumption today and are not included. Even hydropower never exceeded the two percent share it reached in 1955. The importance of hydropower would slightly increase to about a six-percent share in the same year provided that its contribution is converted to primary energy equivalent in terms of the fossil energy needed to generate the same amount of electricity. Dung is not included because of the lack of reliable consumption estimates, although its share in primary energy consumption was considerable in the past. Putnam (1953) suggested that dung, as a fuel source, has had a fairly constant share of 16 percent since the 1860s. This is only a rough estimate, the actual consumption of dung and other agricultural wastes was probably much higher during the last century and may be even higher than ten percent of all commercial energy today. Even the fuelwood consumption time series cast some doubt on their accuracy due to the lack of any fluctuations present in the use of all other primary energy sources.

Global primary energy consumption, according to the five most important sources of energy given in Figure 1.6, indicated the inhomogeneous evolution of the world primary energy system. It is evident that the older forms of energy have been substituted by the newer ones. These dynamic changes are more clearly seen in Figure 1.7 which shows the fractional shares of the total primary energy market taken by these five energy sources. In terms of fractional market shares, fuelwood was already substituted by coal during the last half of the nineteenth century. In 1860, fuelwood supplied about 70 percent of consumed energy, but by the 1900s its share had dwindled to little more than 20 percent. Due to the insignificant use of crude oil and natural gas during the last century, most of the market losses incurred by fuelwood were taken by coal. Consequently, coal's share in total energy supply increased from 30 percent in 1860 to almost 80 percent by the 1900s. By 1910 the rapid increase in coal use ceased, and during the 1920s a phase of decline set in. This decline in the relative share of coal use resembles the market losses of fuelwood fifty years earlier. The replication of this pattern is almost symmetrical because after the 1920s both fuelwood and coal are substituted by still newer sources of energy - crude oil and natural gas.

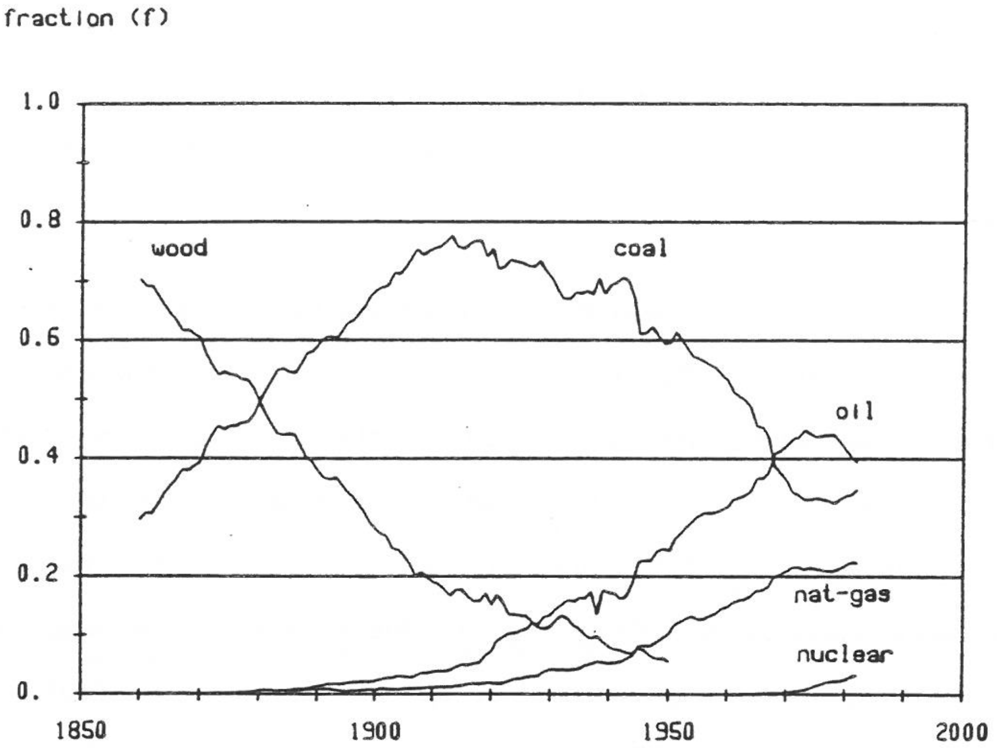


Figure 1.7 Fractional Shares of Major Primary Energy Sources, World.

1.3 Global Energy Substitution

Substitution of the old ways of meeting energy demand by the new ones resembles similar substitution processes on the level of single technologies, such as various steel, paint or turpentine production methods. Consequently, we made a heuristic assumption that energy systems, like other goods and products, can be viewed as technologies competing for a market and that they should behave accordingly (see Marchetti and Nakicenovic, 1979). The evolution in primary energy use, seen as a technological substitution process, is shown in Figure 1.8 on a logarithmic plot of the fractional market shares of the five primary energy sources. The fractional shares (f) are not plotted directly but as the linear transformation of the logistic curve, i.e., $f/(1-f)$ - the ratio of the controlled

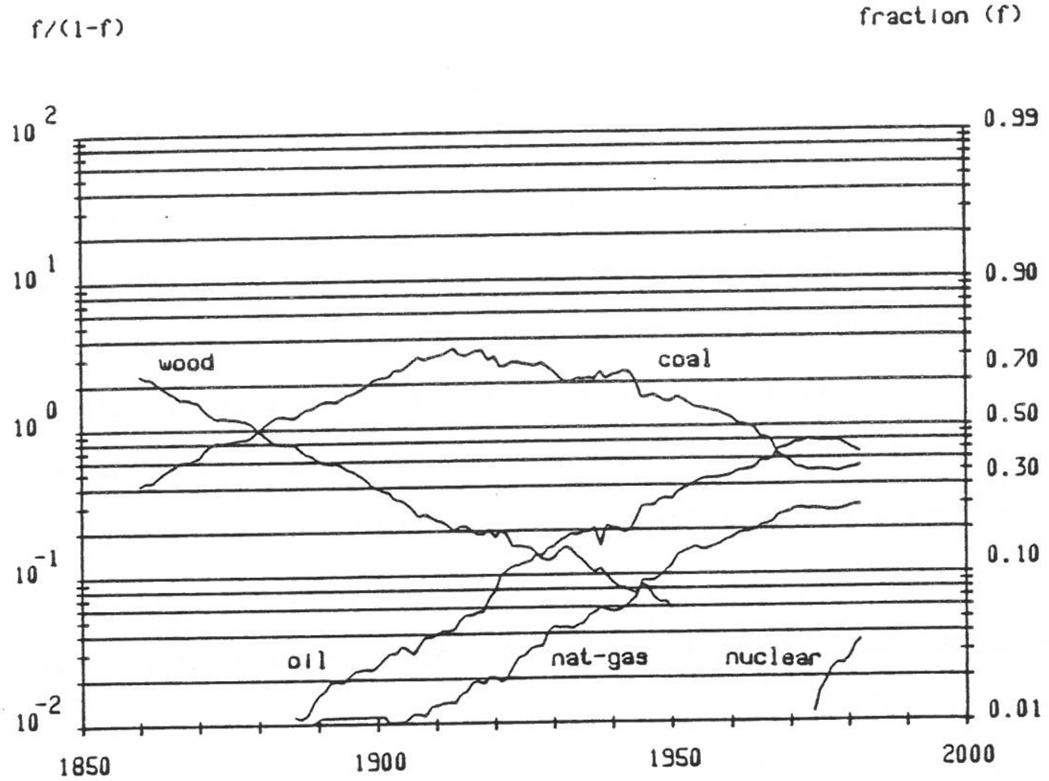


Figure 1.8 Primary Energy Substitution, World.

over the uncontrolled market share. This is the same form of showing a substitution process as that used by Fisher and Pry in their analysis of technological processes. This form of presentation reveals the logistic substitution path as an almost linear secular trend with small annual perturbations. Thus, the presence of some linear trends in Figure 1.8 indicates where the fractional substitution of energy sources follow a logistic curve. It also indicates that energy substitution is similar, at least in this respect, to the technological substitution patterns observed by Fisher and Pry. Figure 1.8 lends strong support to our approach of treating energy sources with the logistic substitution method. However, Figure 1.8 also shows that in the case of multiple competition, logistic substitution is not preserved in all phases of the process. This point is well illustrated by the substitution path of coal, which curves through a maximum from increase to decline in its market shares.

As mentioned above, when dealing with more than two competing technologies, we had to generalize the Fisher and Pry model since the logistic paths are not preserved in all phases of the substitution process. As was illustrated in the substitution path of coal, every technology undergoes three distinct substitution phases: growth, saturation, and decline. The growth phase is similar to the logistic substitution by a new for an old technology in the Fisher and Pry model, but it usually terminates before full takeover of the market is reached. This is followed by the saturation phase which encompasses the slowing of growth and the beginning of decline. During the saturation phase, the secular trend is not logistic. After the saturation phase, the market shares proceed to decline logistically.

In order to describe this more complex substitution process of more than two competing technologies, we have assumed that only one technology is in the saturation phase at any given time, that declining technologies fade away steadily at logistic rates uninfluenced by competition from new technologies, and that new technologies, after entering the market grow at logistic rates. The technology in the saturation phase is left with the residual market share and is forced to follow a nonlogistic path that joins its period of growth to its subsequent period of decline. In other words, the market share of the oldest still-growing technology - for example, coal - can be defined as a complement to one of the sums of the other fractional shares following logistic substitution paths. Thus, during the saturation phase the oldest still-growing technology takes the residual of the market, but eventually its share starts to decline and becomes logistic again. Subsequently, the next newer technology - oil in this example - would undergo the transition from growth to decline. In effect, the technologies that have already entered their period of market phase-out are not influenced by the introduction of new ones. Deadly competition exists between the saturating technology and all other, newer technologies.

This short description of the generalization of the Fisher and Pry model illustrates the most important features of multiple competition that are captured by the logistic substitution model. A more formal presentation of the model is given in the Appendix which translates our assumptions into mathematical language.

Figure 1.9 shows the application of the model to our example of global primary energy substitution. The fractional market shares of the five primary energy sources are plotted on the logarithmic scale as the linear transformation of the logistic function (in the same way as in 1.8). The smooth secular trends are the model estimates of the historical data; the straight lines show where energy sources follow logistic substitution paths. The result is encouraging, since it indicates that this kind of model is capable of reproducing the evolution of energy substitution with very high accuracy. The scatter of the historical data is in good agreement with the smooth trends generated by the model. Significant deviations from the estimated trend can be observed only during relatively short time periods.

The model estimates of the substitution process are extended beyond the historical period up to the year 2050. For such an explorative "look" into the future, additional assumptions are required because potential new competitors such as nuclear and solar energy have not captured sufficient market shares in the past to allow estimation of their penetration rates. Nuclear energy controls slightly more than a two-percent share in all primary energy supply and is therefore visible from the historical data on the graph. Solar energy contributes

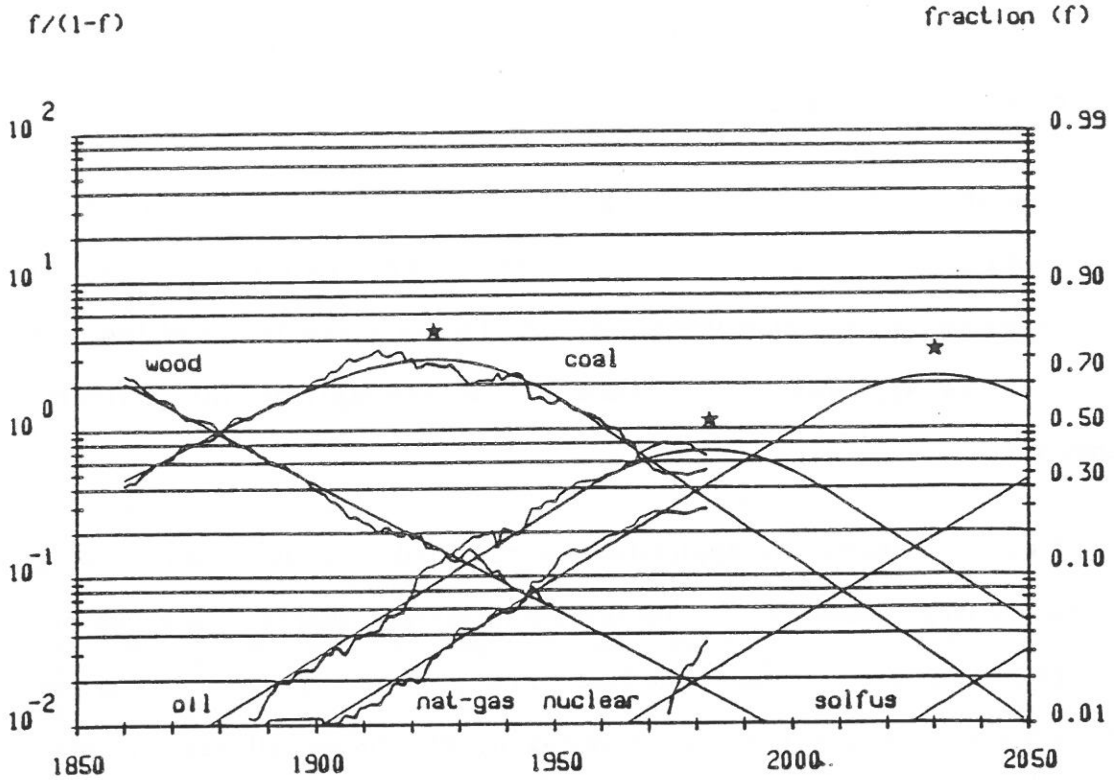


Figure 1.9 Primary Energy Substitution, World (with Projections).

much less than one percent so it is not visible on the graph. Thus, the starting point for market penetration of nuclear energy can be determined from the historical data and dated back to the early 1970s when nuclear power acquired a commercially significant scale with slightly less than one percent share in global primary energy consumption, but its future market penetration rate cannot be determined from the historical data. In order to explore and test the behavior of the logistic substitution model when the competition between the primary energy sources is extended into the future, we made explicit assumptions about the penetration rate of nuclear energy and possible entry of solar or some other advanced source of energy into the market at the one-percent level, as well as its future penetration rate.

In 1982, the net electrical capacity of nuclear power plants connected to the grid (by 31st of December) reached 173 GW(e) (IAEA, 1983). Taking an overall utilization factor of about 55 percent and thermal to electric conversion efficiency of about 34 percent, the nuclear share in primary energy consumption was 3.5 percent. In other words, nuclear energy contributed 282 GW(th)yr/yr to total primary energy consumption of 8162 GWyr/yr. By 1990, according to the IAEA (1983), power plants currently under construction and planned should be in service; thus, the total installed net capacity should be at least 370 GW(e). With the same rough utilization factor of about 55 percent and the total primary energy growth of 2.3 percent per year, this corresponds to about a six-percent share in 1990. We have chosen a more modest nuclear share to account for possible delays in the construction of the planned power plants so that our nuclear scenario prescribes a one-percent share in 1965 and a three-percent share twenty-five years later. For the next energy source, which we symbolically call "solfus" in order to indicate the potential use of both solar and fusion energy, we have made an equivalent scenario with a one-percent share in the year 2025 and three-percent in 2050. In other words, we have assumed that solfus energy will reach a level of commercial utilization by the year 2025 that is comparable to that of nuclear energy in the 1970s.

Figure 1.9 shows these two scenarios and the resulting dynamics of energy substitution throughout the first half of the next century. Note that the penetration trends have almost the same slope for all energy sources including our two scenarios. The extreme regularity and slowness of the substitution process is also conserved in our exploratory use of the model to project the future competition between primary energy sources. We have assumed in the two scenarios that about twenty-five years would be necessary before the new energy sources, nuclear and solfus power, could increase their market shares from one to three percent. Similar penetration rates can be observed for coal, oil and natural gas during the last 120 years. It takes about one hundred years to go from one to fifty percent of the market share, or about fifty years from a ten to fifty percent share. We call this length of time the time constant of the system.

The regularity of the substitution process refers not only to the fact that the penetration rate of various energy sources remains constant over periods of about one hundred years for a given energy source and is almost the same for all energy sources, but also to the fact that most perturbations are reabsorbed after a few years without affecting the long-term trend. Only during the initial phases of market penetration, do the market shares not immediately stabilize to long-term substitution trends. Oil penetrated somewhat faster until it reached a two-percent market share, while natural gas controlled almost a constant one percent of the market for over a decade before both energy sources stabilized to their long-term substitution trends. Our scenario of the future penetration rates for nuclear energy indicates a similar departure of the long-term behavior from the relatively rapid growth before the 1970s.

It is also interesting to note that the saturation levels achieved by all energy sources are much lower than the full market takeover. The introduction of new energy sources and the long time constant lead to maximum market penetrations of between 50 and 70 percent. New energy sources are introduced before the most dominant ones have even reached a 50 percent share. For example, crude oil and natural gas were introduced during the 1880s whereas coal, then the dominant source of energy, reached its maximum market share 30 years later. The stars in Figure 1.8 mark the years 1925, when coal reached its maximum market share, 1980, the maximum share of crude oil and the year 2030, the maximum share of natural gas. Thus, the maxima are roughly spaced at intervals of about 50 years, which corresponds to the time constant of about 50 years for market share increases from 10 to 50 percent. Thus, the dynamics of energy substitution during the last 120 years and over the projected 70 years indicate that the global energy system behaves as though it had a schedule that governs its evolution. In order to further explore the regularity of this process we have used the logistic substitution model to test the robustness of this behavior with respect to different assumptions about the availability and quality of historical data and with respect to different scenarios for the competition of new energy sources.

1.4 Data Base and Scenario Changes

One of the problems in analyzing periods of one hundred or more years lies in the underlying inconsistencies and unavailability of the recorded data. We have tried to check the stability of the estimated parameters and the projections into the future with respect to restrictions in the data base, because the recorded statistics are sometimes unreliable, have gaps lasting for long periods of time, and because they refer to certain energy sources and not to others.

Figure 1.10 shows the robustness of the method to produce stable estimates of the historical substitution with limited input information. The results are encouraging, showing that the relevant information can be extracted from relatively short time periods. The penetration rates of each energy source in the logistic substitution model could be estimated with only two data points (as we have done in the nuclear and solar scenarios), since only two points are needed to define a straight line. Consequently, the large number of statistical data serve to

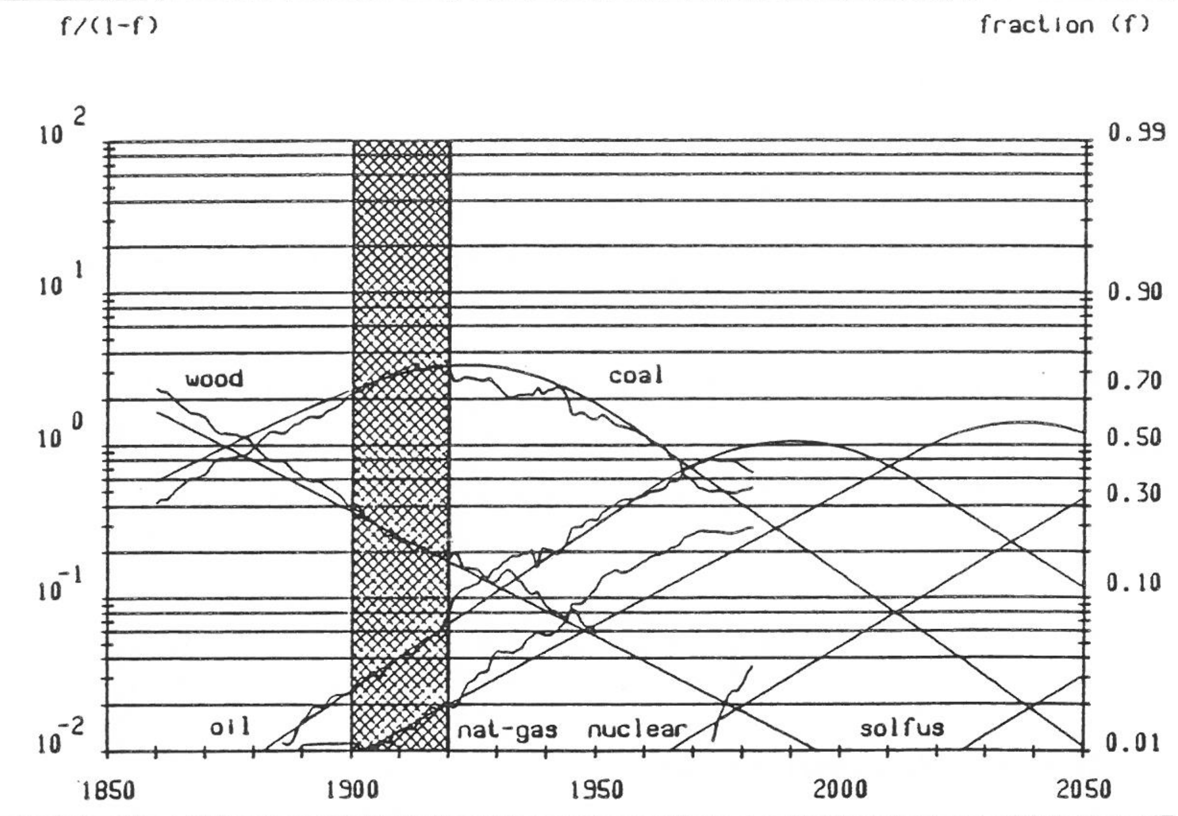


Figure 1.10 Primary Energy Substitution, World (Short Data-Base).

reduce the noise and eliminate fluctuations. Usually twenty years of historical statistics already constitute an excellent data base. Therefore, we have taken a twenty year period from 1900 to 1920 to estimate the model coefficients and "reconstruct" the energy substitution dynamics from 1860 to 1982. This particular period was chosen because it is the earliest that already includes crude oil and natural gas, but it has the disadvantage that gas had reached only a two-percent share by 1920 and consequently its substitution path was still subject to some change.

The smooth curves in Figure 1.10 give the "backward" and "forward" projections for the whole historical period from 1860 to 1982, based on the twenty years of data between 1900 and 1920. The *ex post* prediction shows an extraordinary agreement with both pre-1900 and post-1920 historical development. By superimposing examples with short and full data-base Figure 1.11 shows that natural gas deviates somewhat and results in an error of a few percentage points by 1980. This may seem relatively large, but it is a "prediction" made for 60 years ahead from a small market share in 1920, and with the Great Depression and the Second World War in between. An equivalent projection based on the past twenty years would reach to the year 2040.

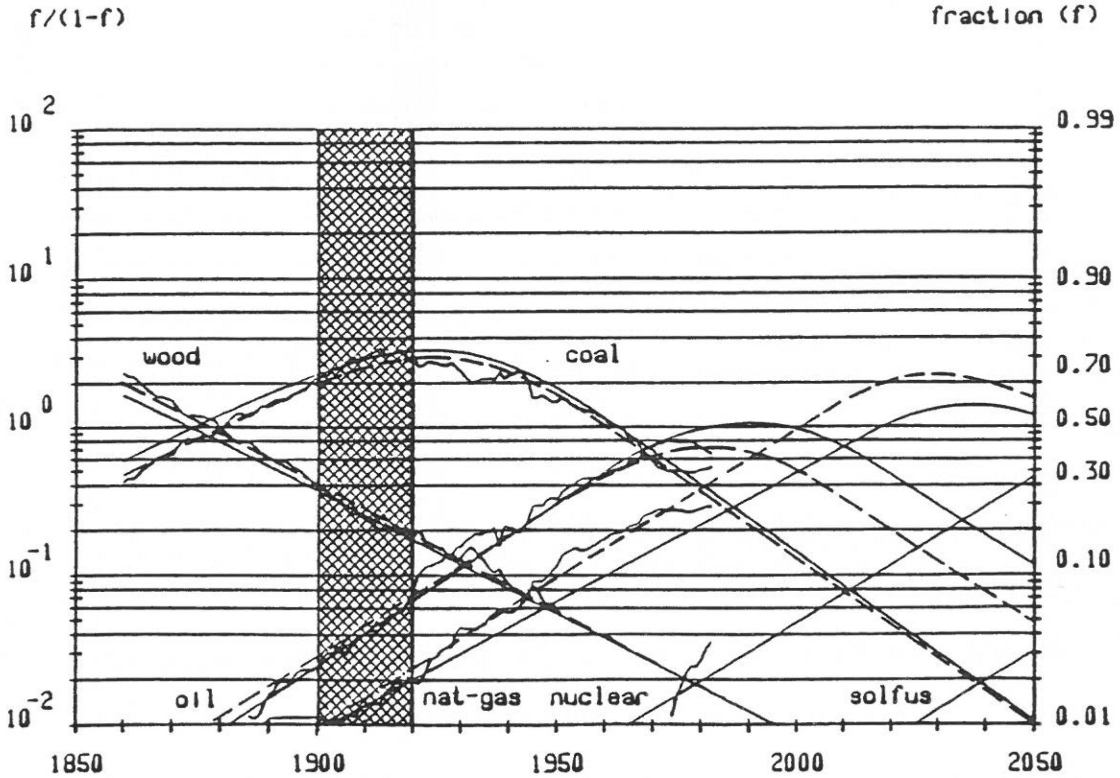


Figure 1.11 Primary Energy Substitution, World (Short and Full Data-Base).

The global energy substitution analysis was based originally on the historical period from 1860 to 1974 (see, Marchetti and Nakicenovic, 1979 and Nakicenovic, 1980). The historical energy substitution with the full data-base given in Figure 1.9 was extended from 1974 to 1982 since in the meantime the statistics for this period became available. Figure 1.12 reproduces the original global energy substitution based on the primary energy consumption data from 1860 to 1974.

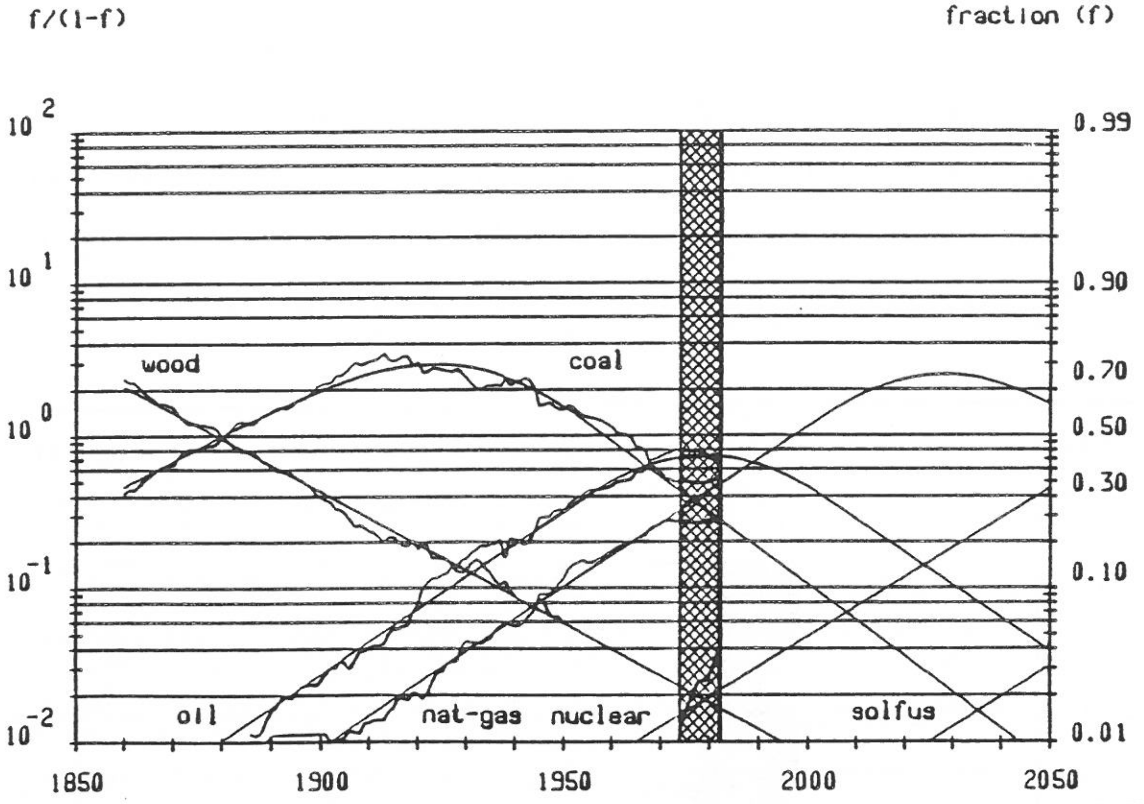


Figure 1.12 Primary Energy Substitution, World (1974-82 Prediction).

Therefore, Figure 1.12 actually gives a real prediction of the substitution dynamics between 1974 and 1982, a prediction for the period of eight years. As can be seen from Figures 1.9 and 1.12, the major difference between the model projections and the actual historical development constitutes small "departures" of the natural gas and coal shares from the long-term trends. It is still an open question whether these "departures" will be reabsorbed during the next ten to twenty years.

The next example, shown in Figure 1.13 illustrates the consequence of the lack of information about the older, now-declining energy sources. We have already mentioned that fuelwood consumption data appear to be too smooth to be accurate when compared with the fluctuations present in the use of all fossil energy sources. In fact, fuelwood data do not represent much more than good

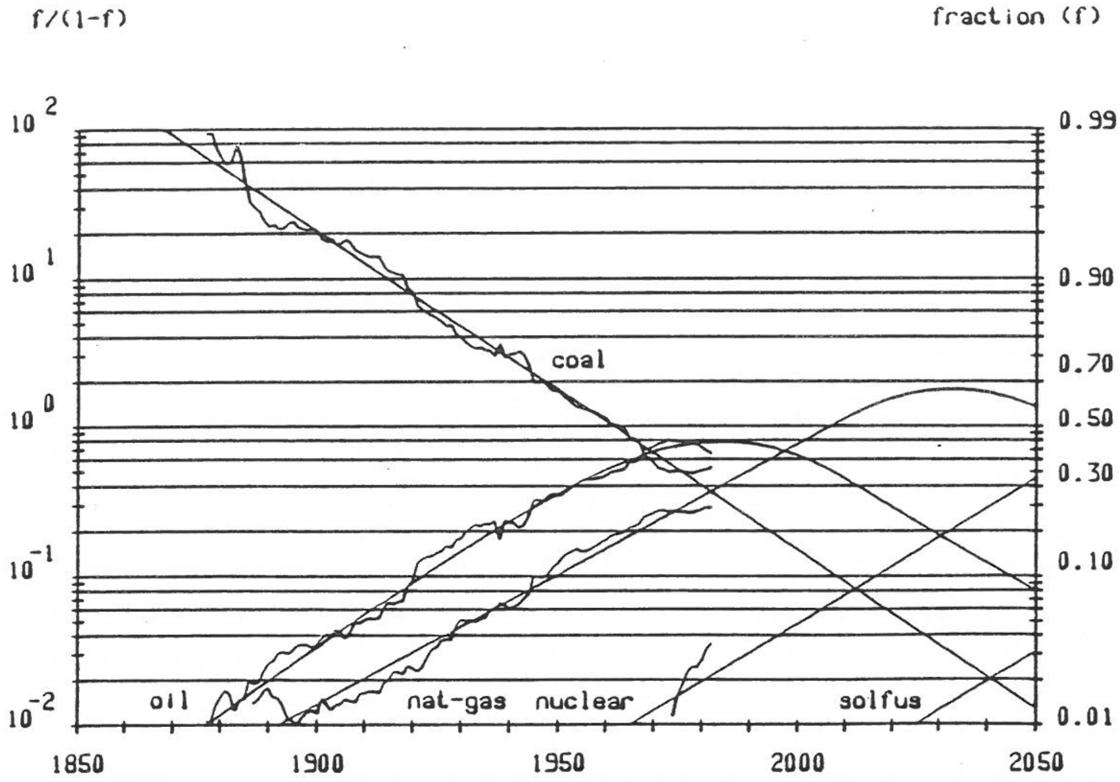


Figure 1.13 Primary Energy Substitution, World (without Fuelwood).

estimates or guesses. Consequently, we omitted fuelwood from our data base and analyzed the competitive behavior of the other primary energy sources that succeeded fuelwood. Figure 1.13 shows that little information is lost from the substitution dynamics by the omission of fuelwood. The differences between the curves with and without fuelwood never exceeded a few percent of the market, as can be seen in Figure 1.14 from the superimposed examples with full and partial information. This example indicates that key information about the dynamics of the market is contained in and can be extracted from the restricted subset of the original data base. This result is not surprising in view of the fact that every market is a submarket. It is, nevertheless, important because it shows that it is not crucial to include older, now-declining technologies in order to extend the substitution dynamics into the future. However, the information loss becomes

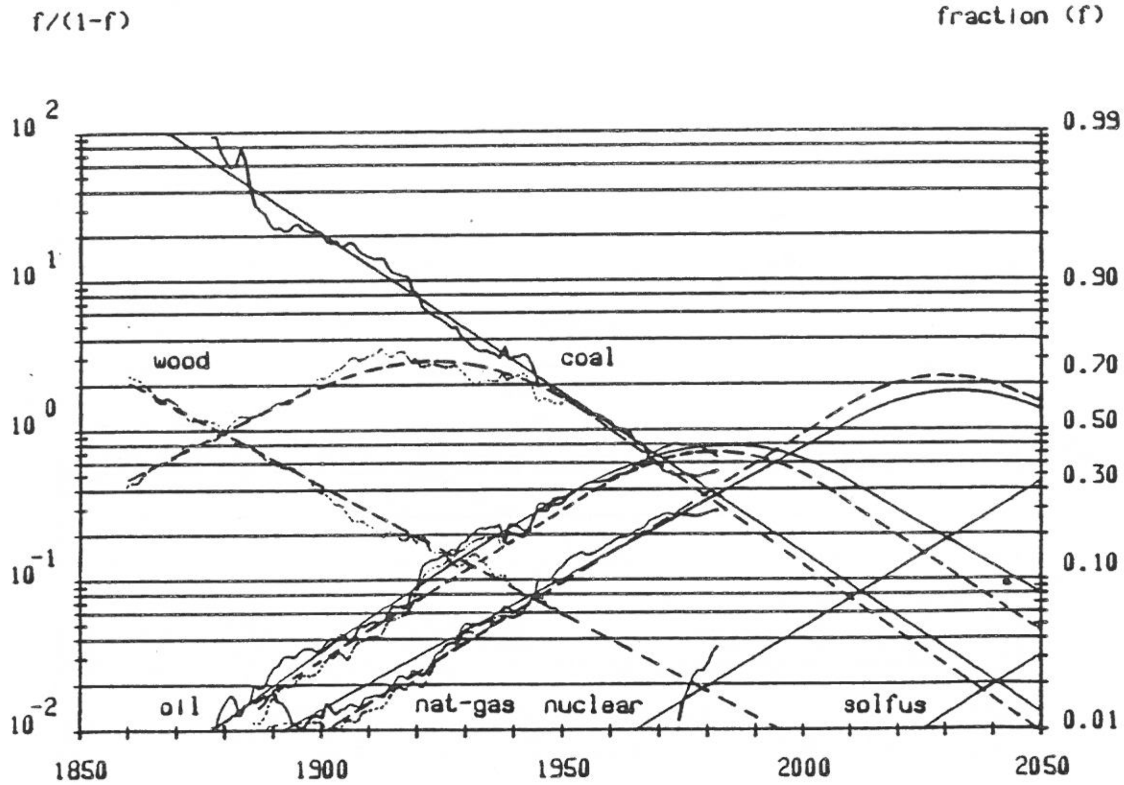


Figure 1.14 Primary Energy Substitution, World (with and without Fuelwood).

significant in the case of "backward" extrapolation to periods prior to the decline of coal, before the late 1940s when fuelwood supplied large shares of primary energy.

The last example indicates the effects caused by changing the scenarios concerning the penetration rate of new competitors. Three different scenarios are given in order to explore the consequences of the uncertainty about the future of nuclear energy. The history of nuclear energy is too short and its market shares are too small to provide a reliable indication of the long term market penetration rates. Figure 1.15 shows the nuclear energy scenario from Figure 1.9 superimposed on the "vigorous nuclear" case with a six-percent market share in 1990 (see, p. 25), and Figure 1.16 shows a similar comparison with the "nuclear moratorium" scenario.

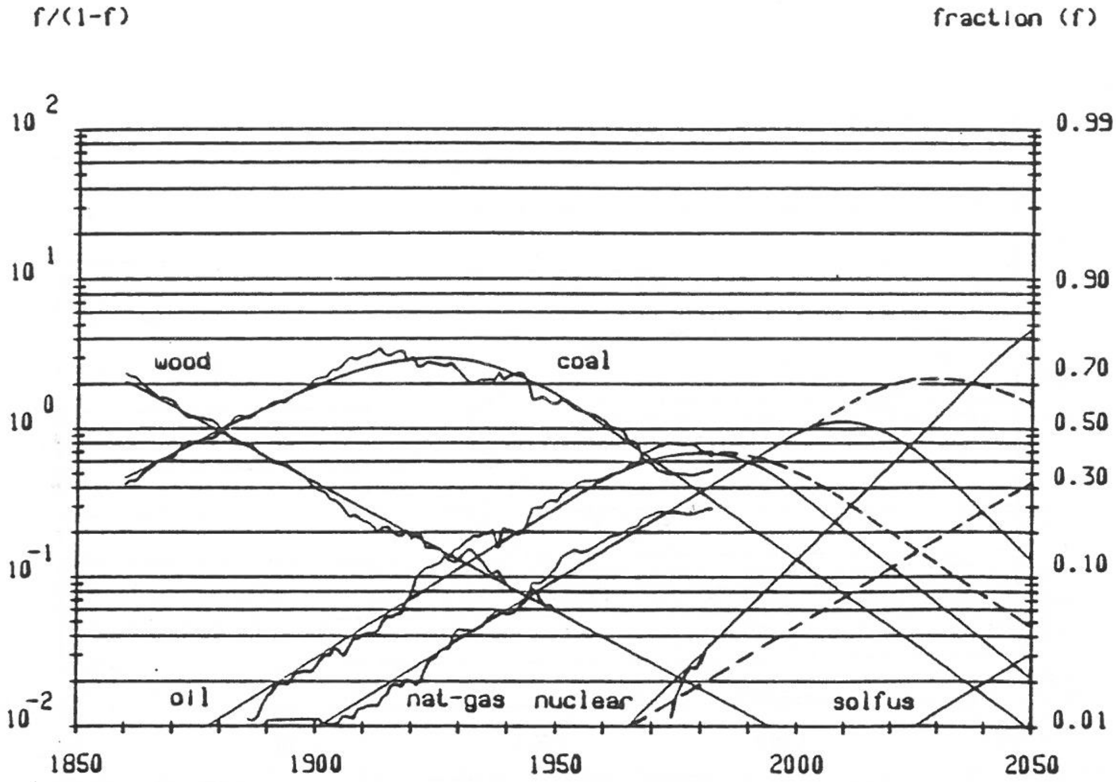


Figure 1.15 Primary Energy Substitution, World (Vigorous Nuclear Scenario).

These examples reveal very interesting properties of the logistic competition. Nuclear energy appears to interact strongly only with natural gas, presumably preempting the markets into which natural gas could have expanded. It reacts only marginally with crude oil, which may cast doubts on the usefulness of installing nuclear power plants as measures for reducing oil imports. Fuelwood and coal, both on their way down, are insensitive to the changes in the assumptions about the "newcomer". Therefore, coal and nuclear energy do not appear as serious competitors in spite of their obvious exchangeability as sources of electricity.

The examples have illustrated the robustness of the logistic substitution process to both changes in the data base and scenarios about new competitors.

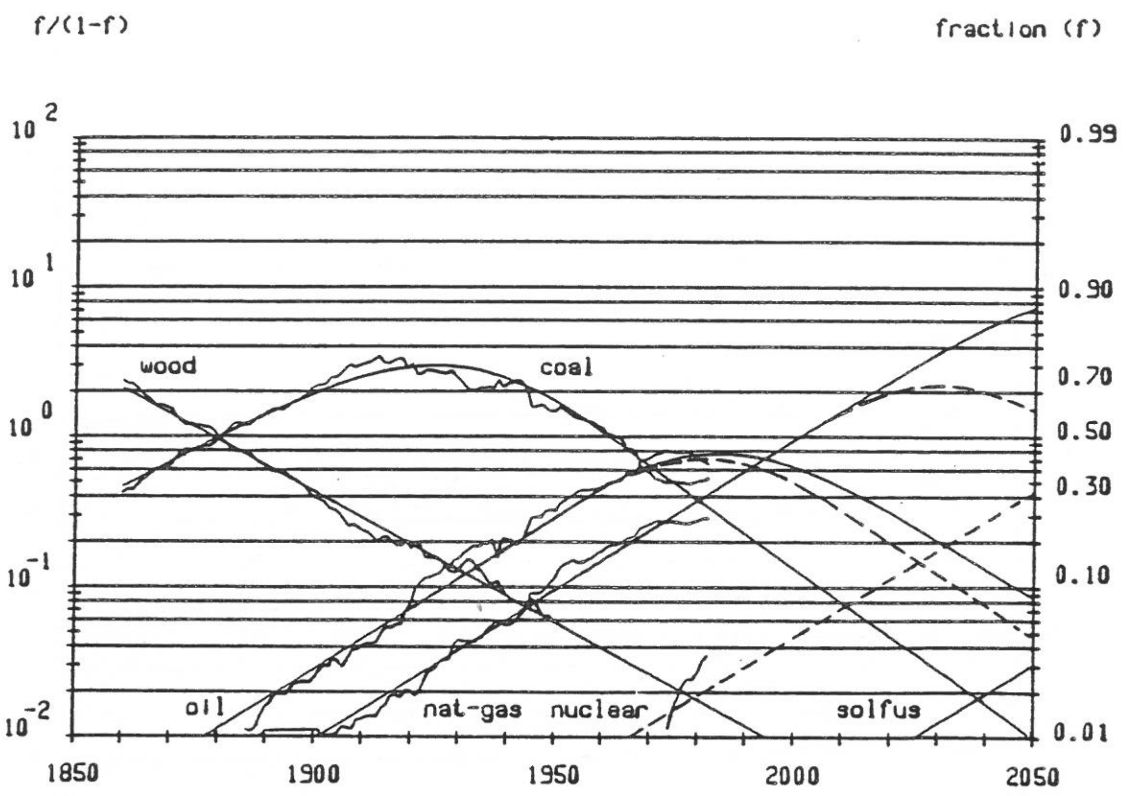


Figure 1.16 Primary Energy Substitution, World (Nuclear Moratorium Scenario).

The unexpected capacity of this approach to "organize" historical data and extract the relevant information about the evolution of energy systems contained in very restricted sets of data is not limited to the example of global energy consumption. To date we have analyzed about 300 examples of technological substitution from the energy field and have discovered that the logistic substitution model offers a consistently good description of the historical data (see, Marchetti and Nakicenovic, 1979; Nakicenovic, 1979 and 1980). Many of these examples were based on historical data ranging back to the 1860s, but the majority of studied cases included historical data going back only two or three decades. This was due to the unavailability of reliable estimates of past energy consumption in many countries and regions of the world. The sensitivity analysis based on global energy substitution illustrated that also examples distilled from short swaths of data provide rather robust information about the future dynamics of energy substitution. They usually have the drawback that the "reconstruction" of the energy system's past behavior is not possible due to the lack of knowledge about energy sources that are no longer used in industrialized countries. Here we will attempt to fill two gaps in the knowledge about the dynamic behavior of technological substitution. First, we will extend energy substitution further into the past and secondly, apply the generalized substitution model to other examples outside the energy field. Most of the examples will be for the United Kingdom and the United States, largely because we had the most success in compiling the sparse data from the last two to four hundred years for these two countries.

2 BRITISH AND AMERICAN EXPERIENCE

The logistic substitution model proved to be capable of describing the competition between energy forms in many different countries and regions. The remarkable regularity of these substitution processes indicates that the invariants can be discovered in the evolution of complex systems over long periods in spite of many fluctuations and numerous changes in the overall social, economic and political environment. In order to investigate the limits of this method of discovering invariants in systems' behavior, we use the logistic substitution model to probe further into the earlier phases of the evolution of energy systems. We will also apply the method to describe the development of the other large infrastructures that are vital in the process of economic development. The examples cover the transportation system and a large industrial sector. We will show how the invariants found in these systems are linked to the innovations that are required to introduce a new competitor on the market, the technological changes that lead to successful market penetration of the new competitor, and the overall structural changes in the economy. These changes are related to the uneven pace of economic growth that is expressed in the long wave phenomenon.

2.1 Pre-Industrial Energy Use

Energy is one of the most important constituents of modern societies. The so-called energy crises of the 1970s, caused by the temporary shortage of crude oil supply and the associated geopolitical, environmental and economic problems after more than two decades of ample and cheap energy supply worldwide, has overshadowed the long history of changes in energy systems, including the numerous problems associated with the supply of sufficient energy for economic growth that provides for improvements in human welfare. Even before the beginning of the industrial age, the provision of adequate energy supply constituted one of the most important commercial and private activities. The evolution of global energy consumption illustrated the slow changes in the structure of the energy system. During the nineteenth century fuelwood was gradually replaced by coal as the world's leading economies entered a period of industrialization. During the last 80 years, coal was itself replaced by new sources of energy - crude oil and natural gas. These global changes in the structure of the energy system are more pronounced in the most dominant economies of the two respective historical periods - the United Kingdom during the eighteenth and nineteenth centuries and the United States during the nineteenth and twentieth centuries. Although fuelwood was the most important source of energy in the past, European countries experienced the serious timber shortages during the nineteenth century that Britain had already experienced during the two preceding centuries. Only the United States were blessed with ample wood supply throughout the period, but even there the large area required for biomass harvesting and expanding agriculture provided serious competition between energy and agricultural land-use. In the United States this need for larger land areas led to further expansion into the Western Territories. However, the consequences of this vast exploitation of timber resources were far-reaching. For example, in 1934 a "75 thousand-dollar (in 1934 dollars) tree-planting project in the middle west has been authorized," ... lending "added proof to the short-sightedness and selfishness of man in the development of the United States. Where once stood mighty forests of virgin timber will now be found vast stretches of denuded territory, swept by burning winds in the summer and stripped of fertile top-soil by the action of rains that run off unimpeded. This is the work of man, who ruthlessly cut down forests for the lumber, without thought for the future or for the effects which might arise from logging operations on an uncontrolled scale." By 1934 "the situation has become so acute as to constitute a national menace" (Scientific American, 1934).

In Europe the problems associated with fuelwood use were grave because the available land area was limited. The larger food requirements needed for the rapidly growing population caused expansions in argricultural activities thereby decreasing the availability of timber. In addition, the increasing pace of industrialization caused new energy needs which posed the new challenge of providing adequate fuelwood supplies. By the end of the eighteenth century the wood crisis was eminent in most of continental Europe, and its alleviation was one of the most important problems during the whole eighteenth century. In fact, fuelwood thefts were one of the most frequent crimes. The first institutionalized measures consisted in more efficient management of forests and in rationalizing and limiting wood consumption to the most strategic sectors of the economy. Other measures were directed toward improving the transportability of wood over longer distances by large canal projects, although wood transport remained to be rather cumbersome and expensive. Finally, many innovations were introduced for improving the efficiency of fuelwood use. They ranged from more efficient stoves in households to more efficient processes of iron smelting, salt extraction and so forth. All of these measures alleviated the shortage temporarily, but with further increases in energy demands the limited supply of wood could not be bypassed.

The ultimate *impasse* out of the wood shortage was solved by the slow substitution of fuelwood by coal. Initially, coal mining practices had to be improved to meet the new challenge, but the biggest obstacle to the widespread use of coal was the requirement of new technologies to use coal effectively in place of wood. The first substitution of wood use, however, did not take place in the energy sector but through new construction methods. Wood was increasingly replaced by bricks and stone in building, but it continued to be the basic material for making ships, furniture, machines and tools until the coal age made large increases in iron and steel production possible.

The eventual introduction of coal as an important source of energy actually turned out to be one of the important motors of the industrial revolution although coal and the associated steam technology did not cause it. Before the advent of large-scale use of steam for industrial processes other sources of shaft power were used such as hydro and wind power and animal and human muscle power. These new uses of coal also helped the transition to a lesser use of fuelwood in its more traditional markets, such as heating and iron production. Railroads emerged as the new transport system based on the use of coal, and they in turn improved the transportability of coal over longer distances and perfected the steam engine. Efficient steam engines made new improvements in coal mining, textiles, and many industrial processes possible.

In order not to present here a very simplistic view of the emergence of the coal era, we should mention that the concept of the industrial revolution is somewhat misleading. The beginning of the industrial revolution is usually dated back to the end of the eighteenth and beginning of the nineteenth centuries. In fact, this period does not really represent a sharp discontinuity from developments in the preceding centuries. The rise of European industry should more properly be regarded as a long evolutionary process dating back to the eighth or ninth century, when aggressive application of water power as the prime mover in manufacture and agriculture, and the use of horses for transport and agriculture as substitutes for oxen and manual labor were initiated.

Cipolla (1976) mentions four main technological developments of the Middle Ages: The diffusion of the water mill starting in the sixth century; the introduction of the heavy plow in the seventh century; the three-field argricultural system starting in the eighth century; and the diffusion of the horseshoe and the new method of harnessing draft animals starting in the ninth century (see also, White, 1972). These technological innovations were not preceded by inventions in the strict sense of the word. The water mill was known to the Romans, but was not used to replace animal and human labor as in the Middle Ages. The horseshoe appears to have been used by the Celts, the heavy plow had a Slavic origin, and the horse harness originated in China. Cipolla points out that the Europeans displayed a remarkable capacity for assimilation between the sixth and the eleventh centuries rather than inventive ingenuity. All of these important innovations allowed larger and more efficient use of available energy sources.

The horseshoe and the new harness increased the effectiveness of using the horse, thereby offering new labor saving opportunities. The ox was increasingly substituted by the horse. For example, in one manor in England (Ramsey Abbey) the number of oxen was halved and that of draft horses quadrupled between 1125 and 1160. The dissemination of iron technology paralleled that of the horse. The amount of iron used in agriculture and machines appears to have been extremely limited before the eleventh century. Later, iron use appears more and more frequently in the records. In the twelfth century, the new ways of using a horse efficiently and the spread of the village smith in Medieval Europe led to the improvement of including metal parts in the heavy plow. Together these innovations made the three-field agricultural practices possible. Fitting new crops into the rotation system expanded agricultural production by reducing fallow land and eliminating the need for the fallow year. Thus, the new way for harnessing animal muscle power and the dissemination of iron processing established the conditions for substantial increases of agricultural productions.

Originally water mills were used for grinding grains, but as cities, trade and manufacturing grew in Europe from the tenth century onwards, the motive power derived from hydraulic energy was applied to an increasing variety of productive processes. The introduction of new mechanisms for vertical movement of hammers, via cams mounted on the axes of the mill, created many new uses for the mills including the fulling of cloth. Beginning in the twelfth century the spread of the water mill revolutionized the textile industry through large labor savings. At the same time mills were also introduced to power hammers and bellows in the production of iron, to power saws for wood cutting, and were even used in the manufacture of paper. The dissemination of the water mill was accelerating throughout the Middle Ages. For example, in late eleventh century England, the areas under Norman rule had 5,624 water mills at more than 3000 locations, amounting to about one mill per 50 households. In the fourteenth century, more than 500 Cistercian monasteries operated water mills and many had five or more units. By the sixteenth century, water wheels were in operation throughout Europe and at some sites the concentration of powered machines was quite comparable to that in factories of the eighteenth and early nineteenth centuries. In 1694, it was estimated that France had 80,000 flour mills, 15,000 industrial mills and 500 iron mills and metallurgical works - a total of 95,000 mills some of them powered by wind. A 1666 survey of the Dnieper River in Russia lists 50 dams and 300 water wheels. One of the largest water power systems was built in the Harz Mountain region of Germany with a network of dams, reservoirs and canals to turn wheels that powered mine pumps, wire-drawing engines, ore-washing and crushing mills, and the bellows of furnaces and forges. The construction of the system started in the 1550s and by 1800 it included 60 dams and reservoirs. The largest dam, built between 1714 and 1721, was 145 meters long and fed water to 225 wheels through a network of 190 kilometers of canals. The aggregate power of the system was estimated at more than one thousand horsepower. For a more detailed treatment of the history of hydraulic power see Reynolds (1984).

Beginning in the thirteenth century, the use of wind mills also spread throughout those areas of Europe with the appropriate geographical and climatical conditions required to harness this power source. Although wind mills never became as widespread as water wheels, they were much more powerful. A single water wheel could deliver anywhere between one and seven horsepower, while the most sophisticated wind mills of the eighteenth century provided as much as 20 to 30 horsepower.

"The proliferation and increasing power of water mills and wind mills, like the increased use of horses, meant more energy for productive uses. Unlike horses, however, the mills supplied inanimate energy. Their widespread use marked the beginning of the breakdown of the traditional world in which man had to depend for power on animal or vegetable sources of energy. It was the distant announcement of the industrial revolution" (Cipolla, 1976).

In spite of these crucial technical innovations of the pre-industrial era, fuelwood remained as the only source of thermal energy in addition to agricultural wastes. It is not surprising therefore that with the growth of population and production, fuelwood became a scarce resource since it was basically non-substitutable as a source of heat and often also as a raw material.

The timber crisis was actually ever present throughout the Middle Ages, starting in the Mediterranean areas already in the twelfth century. By the beginning of the sixteenth century, the southern areas of Europe entered a period of economic decline, so that the demand for fuel and materials stagnated alleviating some of the fuelwood supply problems. But in central Europe, where economic activity expanded rapidly, the shortage of wood was a serious bottleneck to further growth. Especially in England the shortage was already acutely felt in the 1630s. As it happened, instead of destroying the basis for further economic expansion, the energy crisis served to push England on the industrialization road. England was the most successful of all European countries in substituting dwindling wood supplies by coal. Figure 2.1 illustrates the rapid growth of coal use in England by showing the large increases in coal shipments from Newcastle to London between 1655 and 1830.

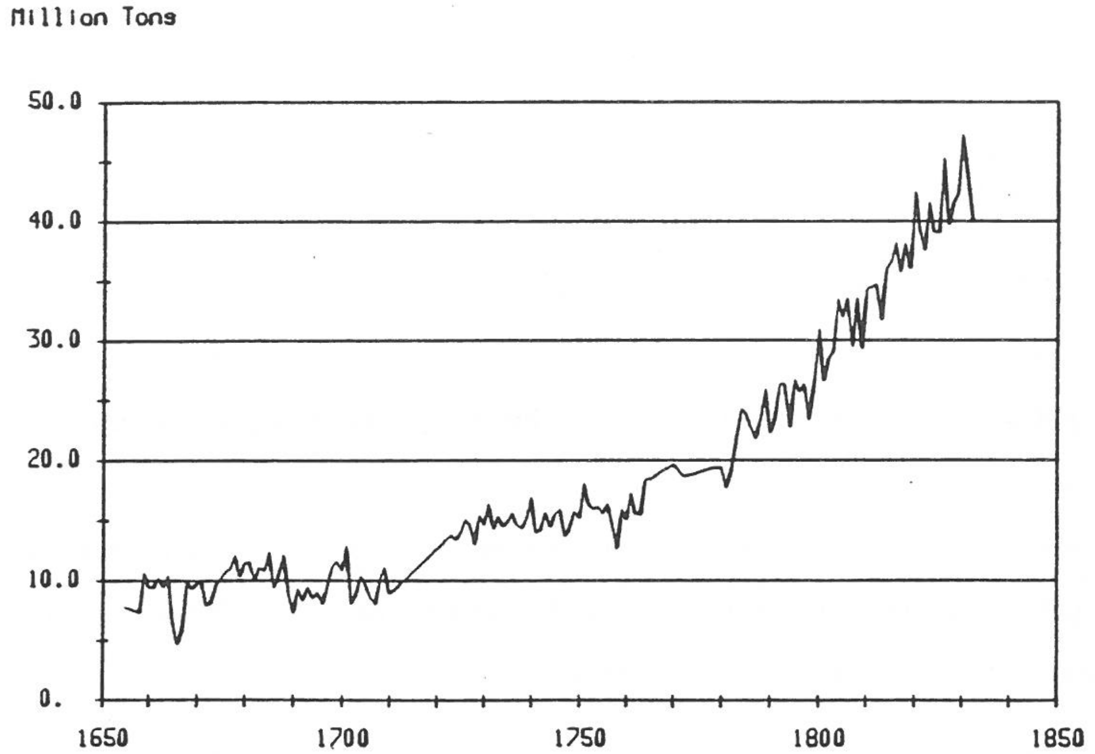


Figure 2.1 Annual Shipments of Coal from Newcastle to London.

2.2 Energy Substitution in the United Kingdom

Initially, most of the available technologies for coal use were rather inefficient. For instance, steam engines were no more than one precent efficient in converting coal into motive power. This low efficiency is comparable to the motive power provided by a horse in terms of the hay necessary to feed it. Also the total horsepower rating of early steam engines was low and no better than could be achieved by water and wind mills. Most of the early mechanized cotton mills utilized only about ten to twenty of the horsepower available from water power. By 1835, the average cotton mill in England utilized less than 35 horsepower (Reynolds, 1984). The replacement of water power by steam engines which began in the 1790s did not represent a large technical improvement because until the nineteenth century the average steam engine output was less than 20 horsepower. The steam engine had the advantage that it provided a more reliable source of power whereas wind and water supply was more uncertain and irregular and necessitated the location of manufacture in windy areas or close to sufficient water supply.

The initial use of coal as a source of thermal energy was also not motivated by the supremacy of coal over other sources of energy, but rather by the acute fuelwood shortage. In fact, up to the early seventeenth century coal was considered technically and environmentally inferior to wood for both industrial and household uses. Once coal started replacing wood in heat generation and once the technologies were developed for its widespread use, it expanded to other areas such as a source of motive power.

An important factor that helped to alleviate the dependency on fuelwood through increased coal use was the rapidly increasing price of timber caused by its imminent scarcity. The large increases in wood prices strongly resemble the oil price increases of the 1970s, which also caused the rationalization of oil use and intensive research for alternative sources of energy. Table 2.1 illustrates the prolonged increases of firewood prices compared with the general commodity price index and price of coal in England between the fifteenth and seventeenth centuries. Although the index numbers represent only rough estimates they indicate that over the two hundred year period commodity and coal price indices increased about three-fold, while the fuelwood index surged by almost a factor of eight (see Cipolla, 1976; and Humphrey and Stanislaw, 1979). Thus, during this period fuelwood became more than twice as costly as coal and other goods. Between the 1620s and 1690s, the price of charcoal doubled while most other prices remained stagnant or even showed a tendency of mild decline (Cipolla,

Table 2.1 Energy and Wholesale Price Indices, England.

	Price Index		
Period	Wholesale	Fuelwood	Coal
1451-1500	100	100	100
1531-1540	105	94	89
1551-1560	132	163	147
1583-1592	189	277	186
1603-1612	251	366	295
1613-1622	257	457	371
1623-1632	282	677	442
1633-1642	291	780	321

1976). Therefore, it is not surprising that attempts to increase coal use were made well before the onset of the industrial revolution. Unfortunately, the data is very sparse about actual energy use during this period in Europe. We can rely only on spot estimates about the use of various energy sources, such as those mentioned about the use of water power and coal shipments to London.

Although the first use of coal as a fuel dates back to the twelfth century in England, the earliest useful annual energy data available are the coal production series starting in 1700. Due to the lack of any comparable time series on fuelwood use in England it is not possible to reconstruct the market penetration of coal and its substitution of wood. Nef (1932) estimates that the wood used as a fuel in 1700 in England was equivalent to about one half million tons of coal, while the coal production in Great Britain already amounted to about three million tons. Fifty years earlier about 215 thousand tons were produced, which corresponds to an annual growth rate of about 5.4 percent (Humphrey and Stanislaw, 1979). In 1700, however, coal could still not be substituted for wood in the smelting of metals. Thus, in spite of large coal production, the timber shortage persisted in iron making. The growth of coal production from the mid-sixteenth century to the late seventeenth century therefore largely represents the substitution by coal for wood as a household fuel.

Figure 2.2 shows coal production and consumption in the United Kingdom from 1700 up to the present. Throughout the eighteenth century coal consumption increased continuously, but not as fast as a century later. Already in 1700 coal became the primary fuel used in the manufacture of alum, copperas, saltpetre, salt, gunpowder, in brewing and so forth, although wood was still the preferred fuel for smelting. During the second half of the eighteenth century, Abraham Darby's technical innovation of using coke in iron smelting opened new markets to coal. Thus, the substitution by coal for wood in industry and the metal trade contributed to the rapid growth of coal use starting in the 1780s and accelerating after the 1820s. At the beginning of the nineteenth century, the age of canals and later of railroads greatly reduced the cost of coal transport and opened many new and rapidly growing markets.

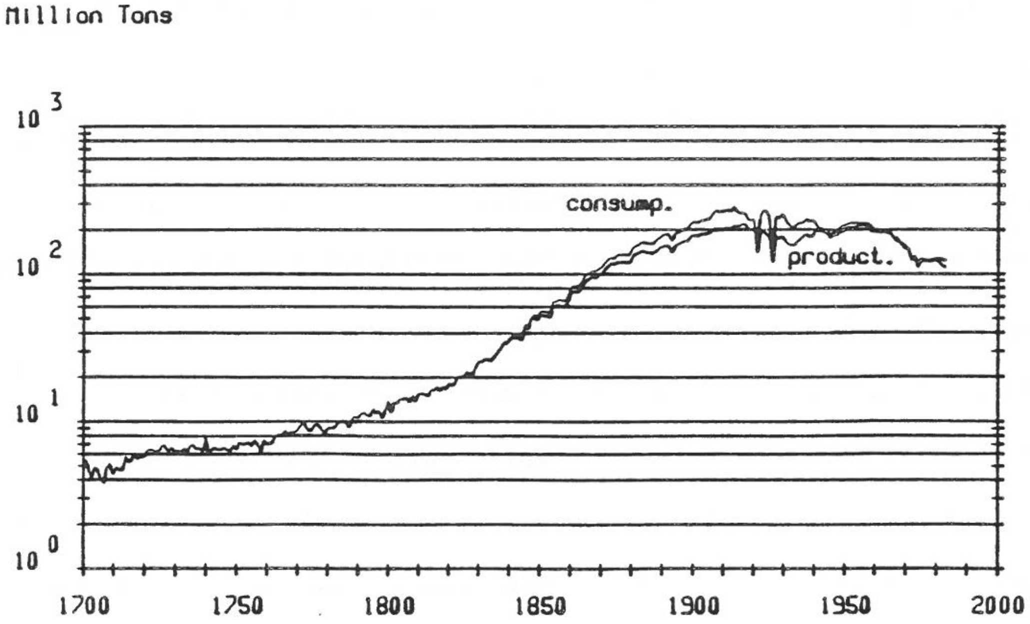


Figure 2.2 Coal Production and Consumption, UK.

During the eighteenth century, wind and water power were still in widespread use in addition to fuelwood. The replacement of wind and water mills began in the 1790s, so that throughout the eighteenth century hydraulic and wind power were the major substitutes for manual labor in providing shaft power. Laxton (1976) estimated that about ten thousand water mills and about two thousand wind mills were operating in eighteenth century England. Humphrey and Stanislaw (1979) showed in a rough calculation that this installed capacity provided an equivalent of about 80 MWyr of shaft power. Considering that early steam engines had a very low efficiency of only a few percent in converting coal to shaft power, the 80 MWyr of shaft power equivalent generated by water wheels and wind mills would translate to about 2.5 GWyr as steam replacement, assuming a three percent efficiency for the average steam engine of the day. In 1760, about five million tons, or less than 5 GWyr, of coal were produced in Great Britain, so that in terms of total primary energy equivalent, wind and water power provided almost one third and coal two thirds. Even if we account for fuelwood use, this rough calculation indicates that already by the end of the eighteenth century coal probably provided more than one half of all primary energy. This rough calculation actually overstates the importance of the traditional energy forms in terms of their fossil energy equivalent, because the actual input of water and wind mills was only 80 GWyr. This energy input was at the time technically and economically not replaceable by steam technologies. Later, when steam machines were further developed, their efficiency also increased. Higher efficiency of converting coal into mechanical energy not only decreases the value, in fossil equivalent terms of the hydraulic and wind mechanical power, but it also illustrates that such calculations are to a degree arbitrary. They only set an upper limit that illustrates the importance of wind and water power in manufacture and agriculture before the industrial revolution. In any case, we can conclude that coal basically became the dominant source of energy by the end of the eighteenth century in England.

Beginning in the nineteenth century, coal finally replaced most of the traditional energy inputs and the growth rate of coal consumption increased paving the way for rapid industrial development. The spectacular growth of cotton textile industries marks the beginning of this process. Later many other industries followed - iron and then steel, railroads, steam ships and so on. The age of coal lasted over two centuries, but in the 1860s crude oil emerged as a new competitor. However, another fifty years went by before oil could capture even a one-percent market share in total primary energy consumption.

Figure 2.3 shows the consumption of all important primary energy sources in the United Kingdom from 1700 up to the present. The figure clearly demonstrates the dominance of coal all the way from 1700 to the 1960s when oil became the most widely used source of primary energy. Crude oil was used for the first time in the 1860s and portrayed extremely rapid growth during the next two decades. From 1900 to 1960, the sustained growth of crude oil consumption was still very rapid, averaging 5.9 percent per year. In spite of such impressive sustained growth, it was not until the 1920s that oil use surpassed the coal consumption levels reached in 1700, almost 220 years earlier. In fact, an acceleration in the growth of new energy sources can be observed. Oil grew more rapidly than coal, and natural gas and nuclear energy, introduced in the 1960s, show still faster growth rates. This large expansion of natural gas and nuclear energy can in part be explained by the preemption of some traditional coal uses. Nuclear energy expanded into the electricity market - a traditional stronghold of

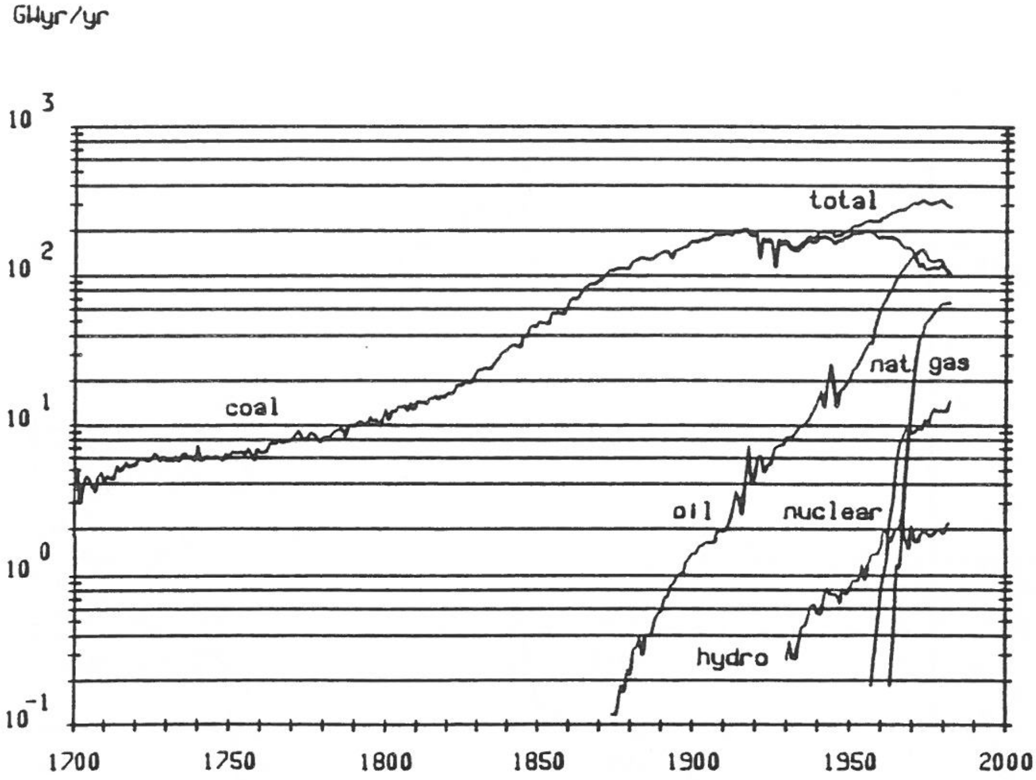


Figure 2.3 Primary Energy Consumption, UK.

coal. Natural gas was in a position to substitute town gas, manufactured traditionally from coal and oil, because a distribution infrastructure was already in place.

Figure 2.4 reproduces the evolution of primary energy use in terms of fractional shares of the four major energy sources together with the logistic trend curves that describe the competition among the sources. The fractional shares (f) are plotted as the linear transformation of the logistic function, $f/(1-f)$, so that the straight lines represent the logistic phases of the competition as it is described by the logistic substitution model. Crude oil achieves a one-percent share of the market in the early 1920s and continued to substitute coal during the next seven decades. It never reached the fifty percent mark because it was in turn substituted by the rapidly growing shares of natural gas. Natural gas

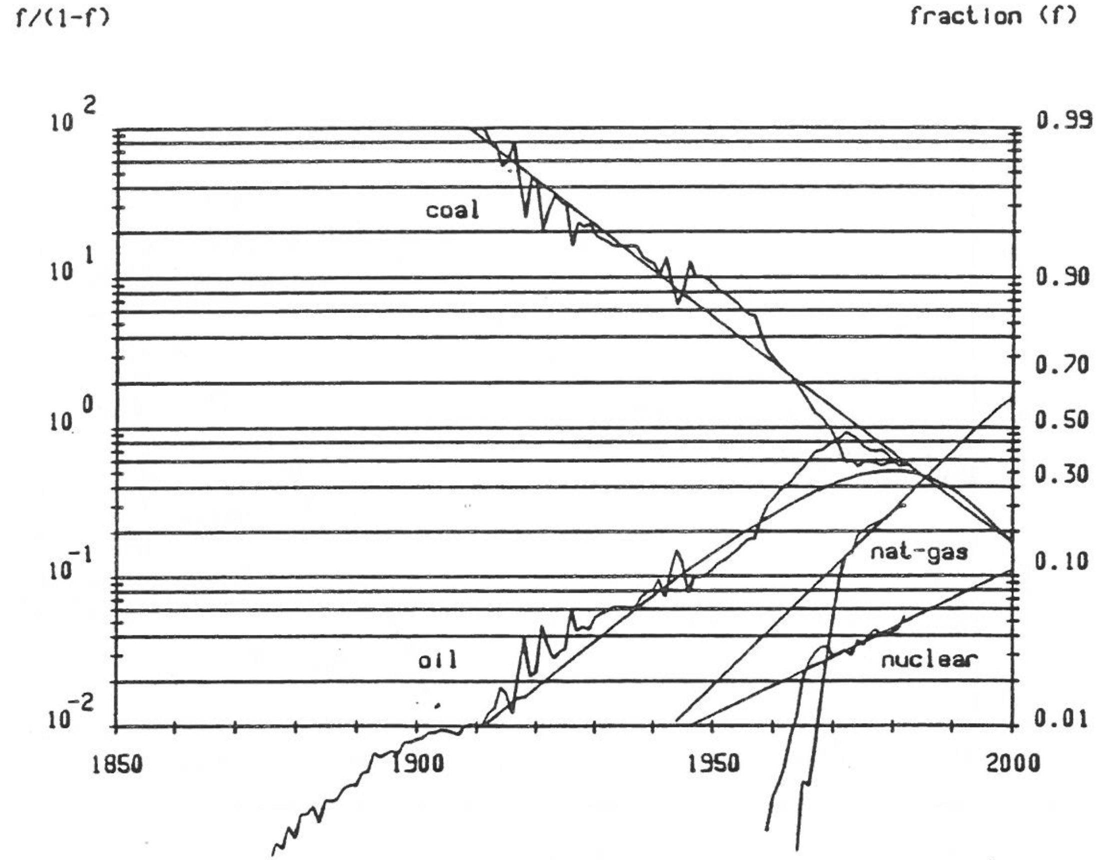


Figure 2.4 Primary Energy Substitution, UK.

actually reached the ten-percent share almost instantly during the early 1960s, transcending the one-percent mark only a few years earlier. We have already mentioned that a possible explanation for this rapid introduction of natural gas could be due to the large gas distribution grid it inherited from town gas. The initial rapid growth of nuclear energy was dampened within a decade when it reached an approximately three-percent market share. Hydropower is not seen in Figure 2.4 because of its marginal use, it never exceeded a one-percent share in total primary energy consumption.

2.3 Energy Substitution in the United States

The records of energy use and production in the United States are very sparse for the Colonial Times. Historical statistics for the period after the Revolution, however, are fortunately almost complete, so that it is possible to reconstruct annual time series for the major primary energy sources. This is particularly interesting because the United States industrialized later than the United Kingdom, therefore offering the possibility to analyze the use of older, now practically extinct forms of energy. Even during the early nineteenth century the United States remained basically an agrarian and rural society. In 1870, farming still contributed almost forty percent of the gross domestic product of the United States (U.S. Department of Commerce, 1970), so that local energy sources constituted an important component of the overall energy supply.

The consumption of mineral energy and fuelwood can be traced back to 1800 when timber certainly represented the most important source of energy. Wood was the principal fuel for domestic as well as industrial purposes. Estimates of the fuelwood use during the last century indicate that it was used lavishly, guaranteeing adequate energy supply. In contrast to Europe and especially the United Kingdom, the United States had vast territories at their disposal, and timber production was usually not limited through resource availability, but rather by the logistics of harvesting and transport opportunities. In practice this means that there was ample supply for local uses. Larger cities in the East were supplied mostly through a growing network of turnpikes. However, water and wind power and work animals also provided substantial inputs to the overall energy supply. As in the Medieval Europe, water and wind power supplied the greater part of inanimate mechanical energy. The rest was provided by animal muscle power and human labor. Especially horses and mules represented a very important source of the motive power needed in agriculture and transport. Even as late as 1920, work animals provided larger aggregate horsepower in farming than tractors and all other agricultural machinery (more than 22 million compared to 21.5 million horsepower, see U.S. Department of Commerce, 1970).

Figure 2.5 shows the primary energy inputs for the United States from 1850 up to the present in five-year intervals. We have included all major energy sources including fuelwood, the energy content of work-animal feed, mechanical wind and water power, coal, crude oil, natural gas, hydropower and nuclear energy. In 1850, fuelwood, animal feed and wind and water mills supplied ninety percent of all primary energy, and coal the other ten percent. Thus, the transition from traditional to commercial energy use was initiated in the United States during the middle of the last century, marking the beginning of the industrial age. In absolute terms, the use of all traditional sources of energy has been declining since 1850 with the exception of slight increases in fuelwood consumption up to the 1870s. This decline continued so that today basically all primary energy use originates from commercial sources and the marginal use of traditional energy forms disappears in the "noise" of the statistical records. During the second half of the last century, coal not only accounted for all of the increases in total energy use, but it also substituted for the losses incurred by traditional energy forms. This explains the very rapid growth of coal during these fifty years with an average annual growth rate of 6.7 percent. During 1870, both crude oil and natural gas entered the energy market, but the enormous expansion of energy use was primarily due to coal until the end of the century. Between 1900 and 1950 the consumption of crude oil and natural gas increased by an average annual growth rate of 8.5 and 6.6 percent, respectively. Overall energy consumption increased from about 100 GWyr in 1850 to over 2400 GWyr in 1981 which corresponds to an average annual growth rate of 2.4 percent.

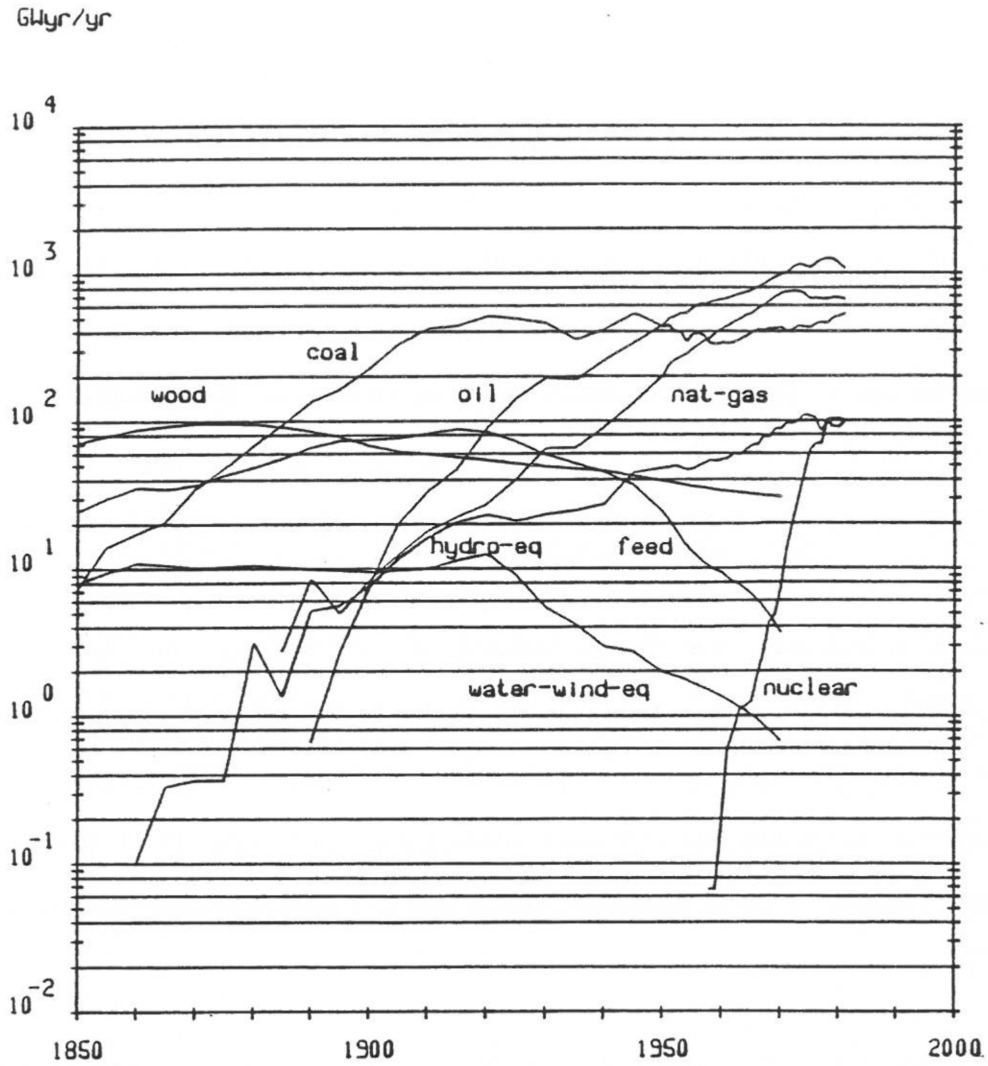


Figure 2.5 Primary Energy (Equivalent) Consumption, US.

Among all traditional energy sources, fuelwood was by far the most important suppling almost seventy percent of all primary energy inputs in 1850. Unfortunately, data on the use of all traditional energy forms during the last century must be taken to represent orders of magnitude rather than precise quantities since they are all based on fragmentary information. Reynolds and Pierson (1942) based their estimates of fuelwood consumption on the population size and distribution, climate, housing conditions, and the availability of wood in the various regions of the United States. Historical records on actual wood use are sparse. As an explanation of this fact, Reynolds and Pierson remarked that there was probably no need to "write about firewood, or even record statistics about it" since "cordwood was about as plentiful as air" in the United States and "nobody wrote about air" use either. Fisher (1974) estimated the feed energy content of work animals by multiplying the number of farm and nonfarm horses, mules and oxen in use with the annual average energy content of the feed required by them. This calculation is based on an annual energy requirement of 3 kWyr per animal, derived from an average animal weight of about 750 kilograms, an average daily consumption of 1 kilogram of feed per 50 kilograms of animal weight, and an average energy content of about 0.5 Wyr per kilogram of feed. Computed in this way, the energy content of work animal feed was the second most important traditional energy source in the United States. It is interesting to note here that although oxen constituted almost 30 percent of all work animals in 1850, horses and mules had displaced them all by the 1900s. Thus, we have here another example of technological change during the early period of American economic development.

Wind and water mills and sailing vessels constitute the last traditional energy source accounted for in our statistical records. An appropriate treatment of this traditional energy source is the most intricate of all energy forms. The difficulty arises from the fact that it can be accounted for either in terms of the direct mechanical energy (inputs) provided, or in terms of equivalent amounts of other energy forms required to produce the same mechanical work. During the last century only animal and human work could have been substituted for the mechanical power provided by wind and water flow. These two possible accounting methods are equivalent to the two alternatives of calculating hydropower consumption either in terms of energy inputs (i.e., amount of electricity generated, sometimes called primary electricity) or in terms of fossil energy requirements to produce the same amount of electricity. In Figure 2.5 both mechanical water and wind power and hydropower are given in terms of energy equivalents. Water and wind consumption are calculated in terms of the feed equivalent energy that would have been required by work animals to generate the same mechanical energy. Hydropower is given in terms of the fossil energy needed to generate the equivalent amount of electricity at the prevailing average efficiency of power plants in corresponding years. This calculation method has the disadvantage of overstating the importance of these two energy forms, especially the contribution of wind and water power. The average efficiency of work animals in converting the energy content of feed into mechanical work is very low and does not exceed four percent. This means that the mechanical energy of wind and water power has to be multiplied by at least a factor of 25 to obtain the feed equivalent energy. In addition, it is simply unrealistic to calculate total energy inputs to the American economy by implicitly assuming that all wind and water mills could have been substituted by horses and mules had it been required. To replace wind and water power in 1860, for example, would imply the feasibility of increasing the 12 million work animals by 30 percent. Furthermore, with the onset of the coal age wind and water power were not replaced by work animals but rather by steam. As in Figure 2.5, Figure 2.6 shows the primary energy inputs of the United States with the difference that the direct mechanical energy of wind, water and hydropower is given in terms of electricity inputs. This does not affect the overall pattern of energy use, since both of these energy sources provided marginal energy inputs that even in Figure 2.5 never exceeded more than a ten-percent share.

In the case of the United States we have the unique opportunity to analyze the substitution of traditional energy forms by commercial energy sources. This

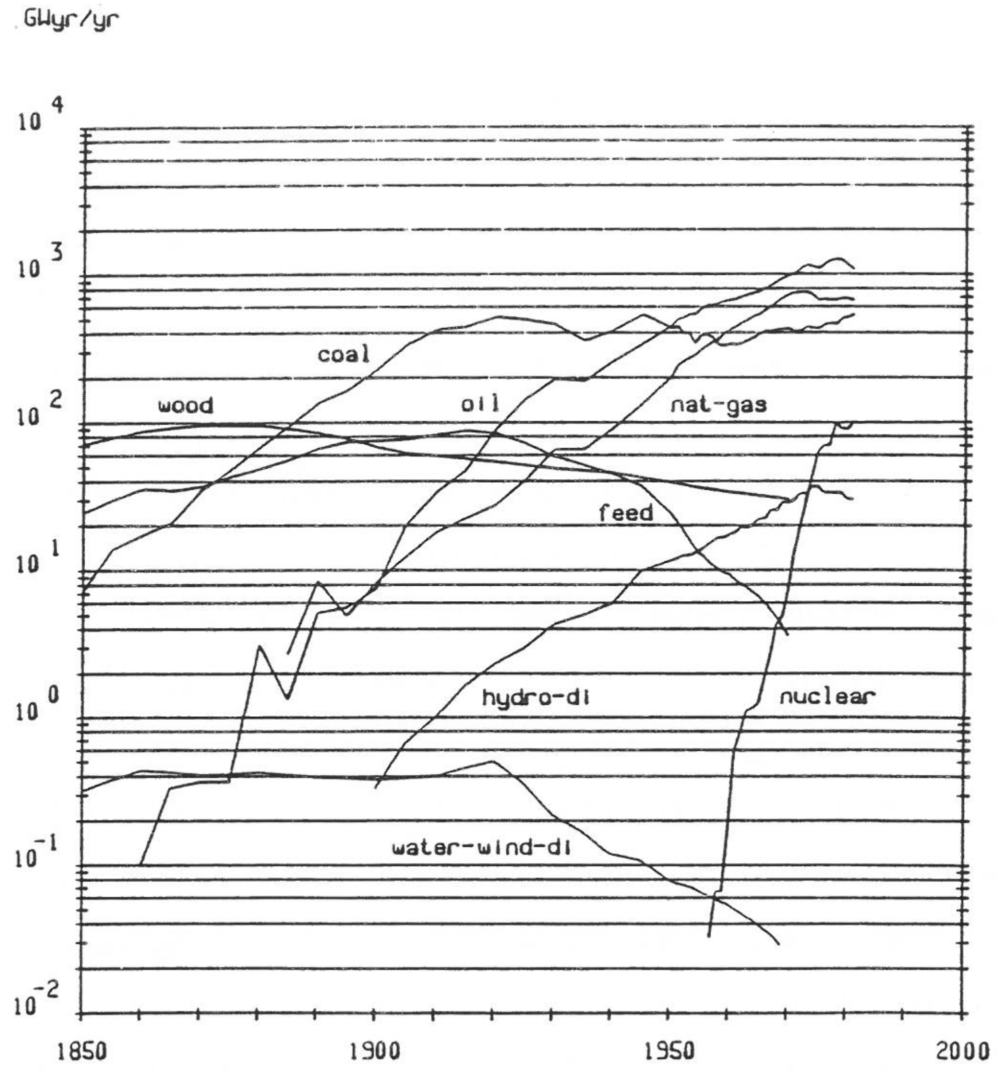


Figure 2.6 Primary Energy Consumption, US.

substitution process can almost be considered as the "proxy" indicator for the pace of industrialization and the economic structural change from agricultural to industrial production. The application of the logistic model to describe this process is shown in Figure 2.7 together with the historical shares of the two broad classes of energy sources. We have grouped fuelwood, animal work, and wind and water mills under the umbrella of traditional energy sources. They all represent renewable energy forms that were basically only suitable for local use by a rural

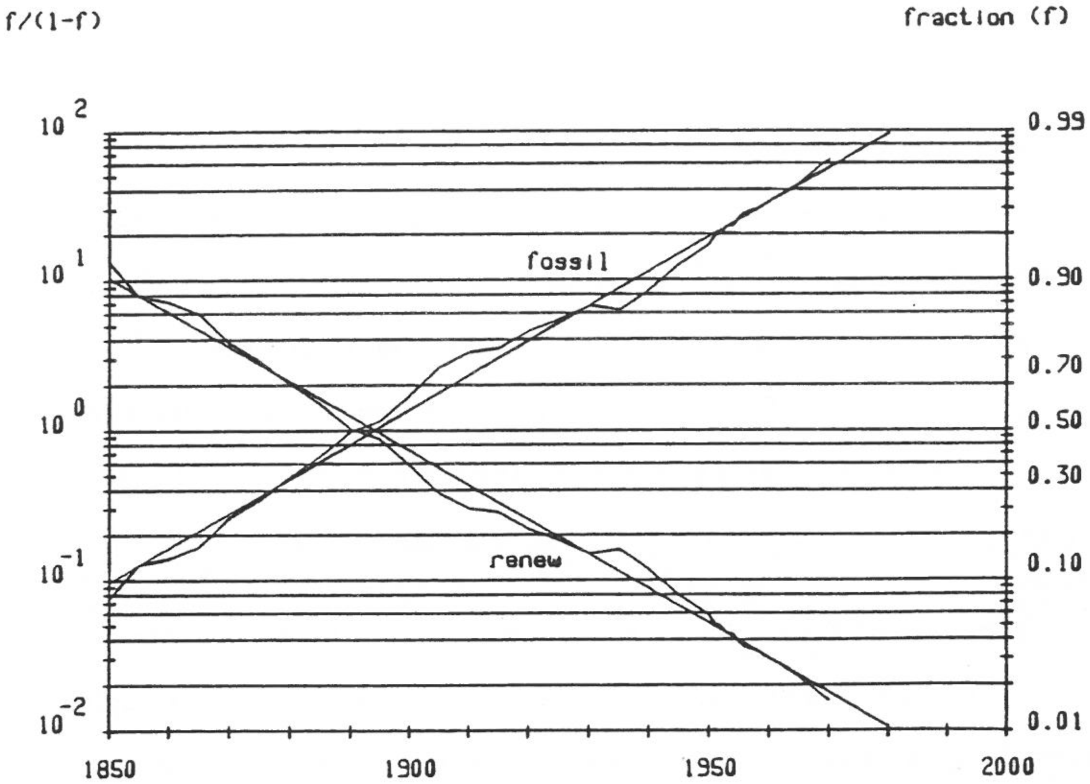


Figure 2.7 Commercial-Fossil for Traditional-Renewable Energy Substitution, US.

society with small concentrations of industrial production in a few urban areas. The traditional energy forms are shown in competition with commercial energy forms including coal, crude oil and natural gas. Until the 1900s almost all fossil energy consumption was based on coal use. The plot shows the two classes of energy use in terms of their respective fractional market shares (f) of total primary energy use as the linear transform of the logistic function, $f/(1-f)$. The substitution process is remarkably regular over the whole time period of over 130 years. It is interesting to note that the fifty-percent mark in the substitution of commercial for traditional energy was reached shortly before the turn of the century. In the United Kingdom this mark must have been achieved at least one hundred years earlier, although this represents only a rough estimate as we have argued in the preceding section. This difference in the shift from traditional to commercial energy use roughly indicates the temporal difference of the industrialization process in the two leading economies during the industrial age. If we extrapolate this energy substitution process in the United States back into the past, the emergence of coal (commercial energy) dates back to the 1820s. In fact, we will show later that this is a very accurate estimate, indicating once more the remarkable regularity of the substitution process. The time constant is quite long - more than 80 years were required before commercial energy sources could capture fifty percent of the market. The corollary of this observation is that traditional energy sources also sustained their decline for over 80 years from the fifty-percent share to the one-percent mark in 1980.

Figure 2.8 and Figure 2.9 show this substitution process from the perspective of individual energy sources. The difference between the two examples is that Figure 2.8 gives water and wind power in terms of feed energy equivalent, and hydropower in terms of fossil energy equivalent and that Figure 2.9 gives the actual energy inputs of these two sources. Thus, Figure 2.8 overemphasizes the role of these two energy sources. This does not affect the structure of the primary energy substitution process however, except for small shifts in the market shares of other energy sources. In both examples the logistic substitution model describes the substitution paths with remarkable accuracy. The departures of historical market shares from their long term paths last for over two decades only to return to the trend after the prolonged perturbation. This is the case with the market shares of coal and oil during the 1940s and 1950s, and fuelwood and animal feed during the 1860s and 1870s. It may also indicate a possible reabsorbtion of the departure of coal and natural gas market shares from their long term paths during the last ten years. The substitution process clearly indicates the dominance of coal as the major energy source between the 1870s and 1950s. The last phases of railroad expansion during the first 30 years of

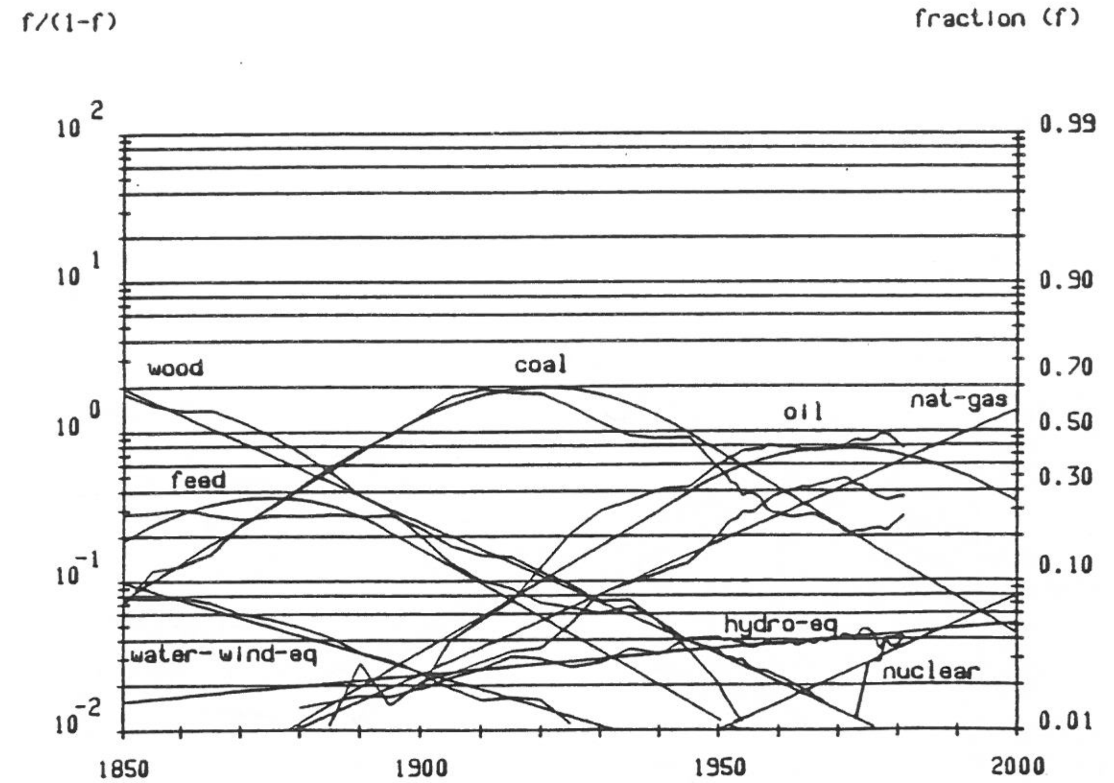


Figure 2.8 Primary Energy (Equivalent) Substitution, US.

this period and the growth of steel, steam ships and many other sectors are also associated with and based on the technological opportunities offered by the mature coal economy. After the 1940s, oil assumed the dominant role in parallel with the maturity of the automobile, petrochemicals and many other modern industries.

The evolution of commercial energy use in the United States has a longer recorded history than the use of traditional energy sources. Figure 2.10 gives the annual consumption of all commercial energy sources and fuelwood starting in 1800. Here again we have two possible representations of hydropower and they are both given in Figure 2.10, the direct energy inputs and the fossil energy equivalent, respectively. The fossil-equivalent of hydropower tends to overemphasize the actual contribution especially during the first few decades of

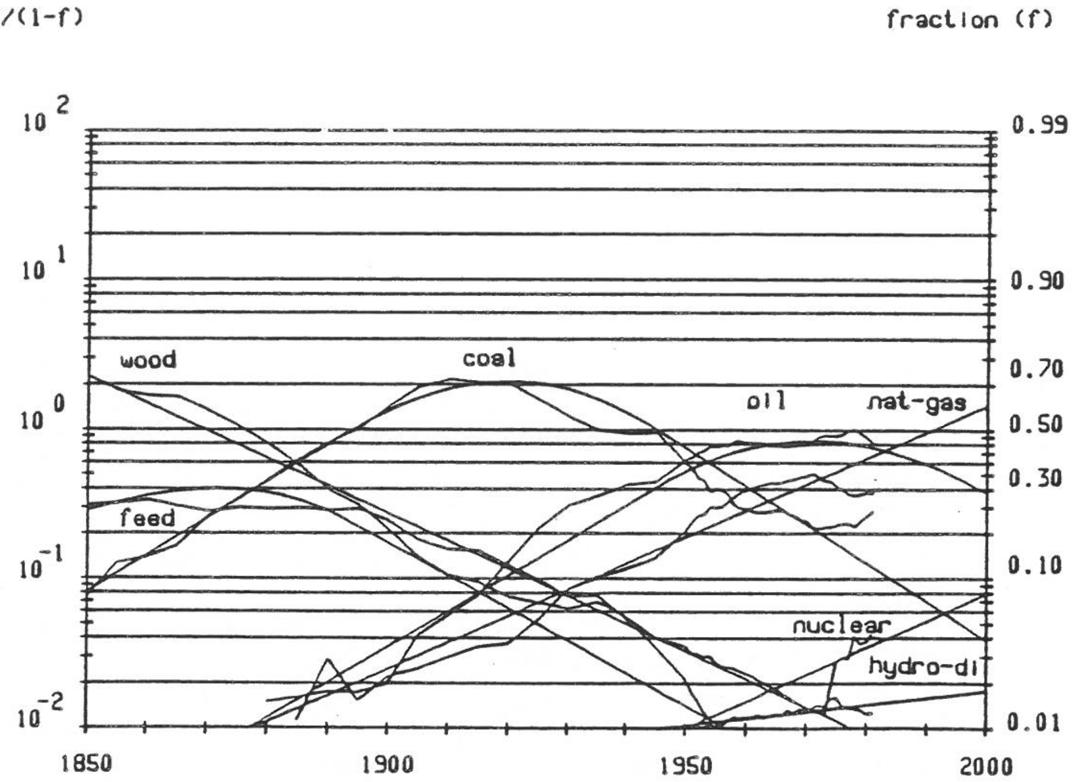


Figure 2.9 Primary Energy Substitution, US.

the twentieth century because the prevailing efficiencies of coal to electricity conversion were very low at that time. For example, in 1920 the average efficiency of installed power plants did not exceed ten percent compared with over thirty percent in 1980. In any case, hydropower shares of the primary energy inputs were not very large, reaching slightly more than four percent in terms of fossil equivalent or little more than one percent in terms of direct energy inputs during the 1970s.

Figure 2.11 shows the substitution of the five most important commercial sources of energy and fuelwood. The logistic substitution model in this example also describes with high precision the evolution of primary energy consumption in the United States. Due to the dominance of fuelwood as the major source of energy during most of the last century, the information loss associated with the

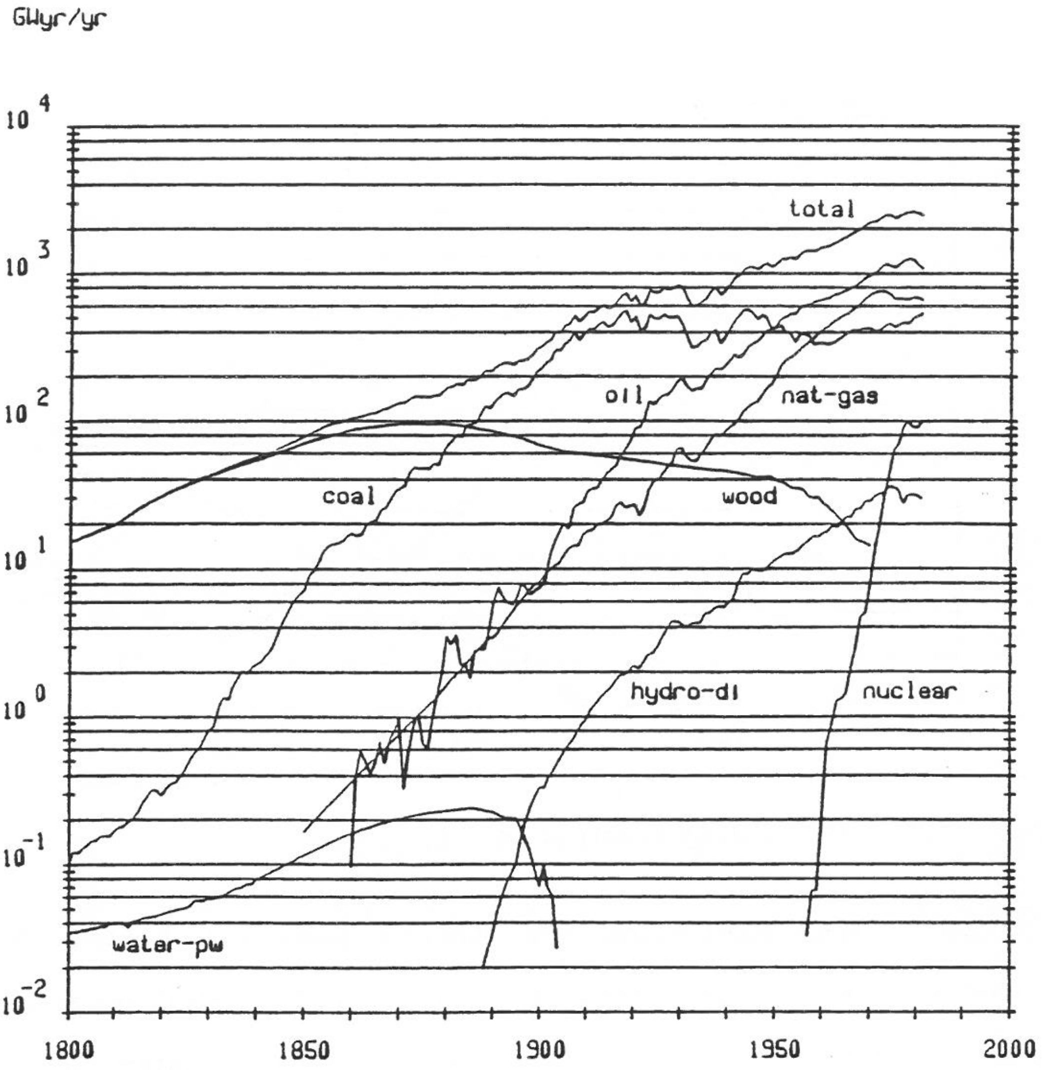


Figure 2.10 Primary Energy Consumption, US.

lack of adequate annual estimates for animate mechanical inputs to the energy system is not very large. Direct wind and water power are included in the data set, but due to their low contribution to total energy supply, when expressed in terms of their actual energy input, they are not observable at the one-percent level. Thus, before the 1820s fuelwood provided for virtually all the energy needs of the United States. Coal entered the competition process in 1817, which corresponds almost exactly to our extrapolation based on the previous example

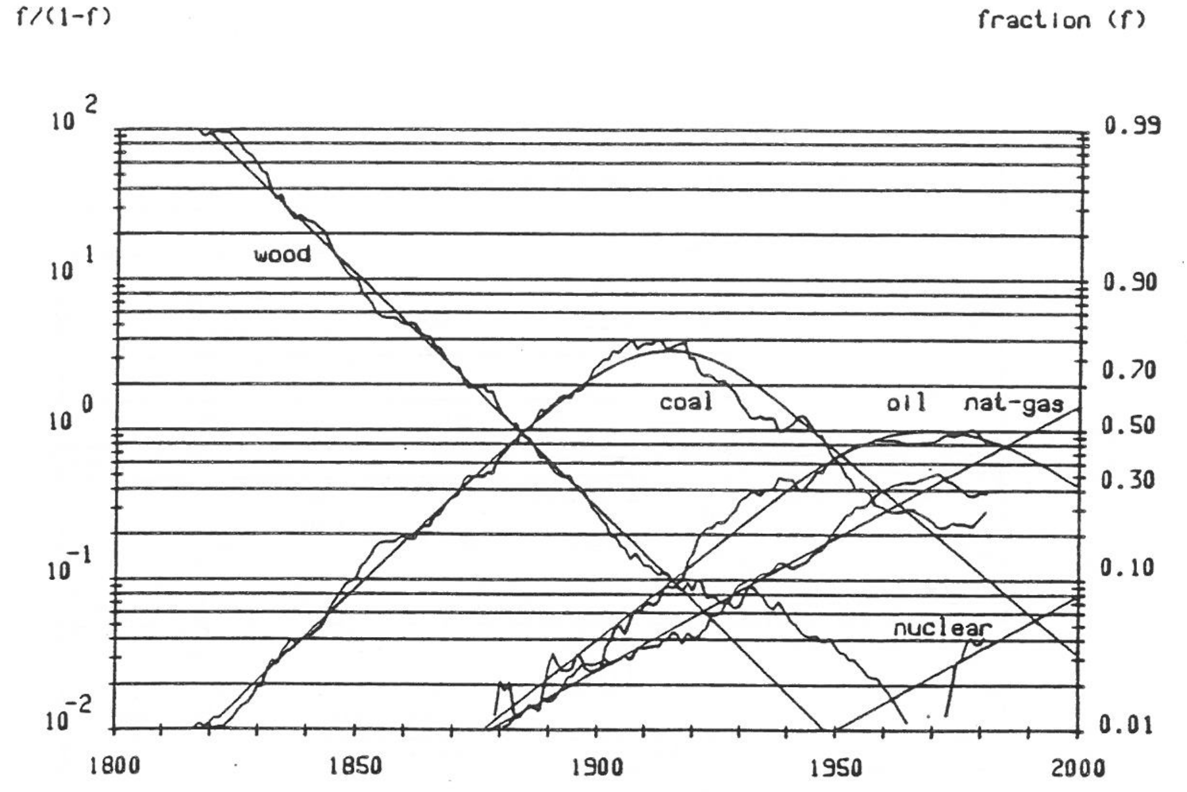


Figure 2.11 Primary Energy Substitution, US.

of primary energy substitution. This example illustrates the senescence of fuelwood and the rise of coal very clearly. Up to the late 1880s it was essentially a two technology market - whatever gains coal made were translated into losses for fuelwood. The initially slow introduction of crude oil and natural gas during this decade translated into market dominance 80 years later. The dominance of fuelwood and coal show an interesting symmetry, each period of dominance lasting slightly over 60 years.

The prolonged use of fuelwood in the United States compared with Europe and especially Great Britain, shows that wood was an important source of heat and power for early industrial purposes well into the second half of the nineteenth century. In the United States the steam age already began in the economy based on wood use. The first steam-boats and locomotives were fired with wood, which remained the principal fuel used by railroads until about 1870 (Shurr and Natschert, 1960). The only other large use of wood was found in the iron industry. Around 1850, more than half of all the iron produced was still smelted with charcoal (see, Figure 2.12). Nevertheless, during this early period of industrialization, the United States was still basically a rural society, so that the total amount of fuelwood consumed in manufacturing and transportation was small compared to the huge quantities used in households. In 1880, the domestic use of fuelwood still accounted for more than 96 percent of fuelwood consumed (Shurr and Natschert, 1960). At the same time, however, coal already supplied almost one-half of all energy needs, most of it being used by emerging industries. In 1880, coal supplied almost ninety percent of the fuel used for smelting iron (see Figure 2.12). Thus, the end of the last century marks the beginning of the industrial development period in the United States.

In the United States the first use of crude oil and natural gas dates back to 1859. During the 1880s, both of these two energy sources reached the one-percent market share. From this point on the use of crude oil expanded somewhat faster as time progressed. In 1950, crude oil consumption surpassed that of coal, and natural gas use surpassed coal nine years later. It should be noted that as late as the 1920s the use of crude oil was not much larger than the consumption of fuelwood. It is remarkable that the structure of energy consumption changed more during the period of oil dominance when compared with earlier periods. The 1950s, when oil bacame the dominant source of energy, represent the beginning of more intense competition between various energy sources both in the United Kingdom and the United States. Until this period coal had essentially been the only important source of energy in the United Kingdom for almost two centuries. In the United States this role was held initially by fuelwood and after the 1890s by coal. All the way up to the 1950s, the energy source that dominated the energy supply at the time also contributed more than one-half of all primary energy consumption. After the 1950s in both countries each primary energy source contributed less than one half of primary energy. In both countries crude oil was close to achieving a fifty percent share during the 1970s, but before actually surpassing this mark proceeded to decline. Thus, during the last three decades three important sources of energy shared the market with no single source having a pronounced dominance, which is contrary to the pattern observed during earlier periods.

The logistic substitution model indicates that it is possible to describe the broad features of the evolution of the energy system in the United Kingdom and the United States over very long periods of time by rater simple mechanisms, in spite of so many turbulent and profound changes since the beginning of the industrial revolution. We also applied the model to describe energy substitution in some three hundred examples ranging from primary to final energy use, electricity generation, etc. (see, Marchetti and Nakicenovic, 1979; Nakicenovic, 1980). Thus, it is evident that the logistic substitution model is a powerful tool when applied to technological change within the energy system. The changes within the energy system are easier to record in terms of long time series since a natural common denominator is available for measuring the contribution of each important component of the system - the contributions of all energy sources can be measured in common energy units. Unfortunately, such a relatively simple and common unit is not available for describing the evolution of other systems. To describe changes in other sectors of the economy, only one obvious common unit is available - the monetary value of the various technologies and activities of the sector. This is not however an appropriate measurement unit for very long time periods, since the price system itself changes with the structural changes of the whole economy. The energy content of a ton of coal depends only on the quality of coal and it is independent of the time period when the coal is mined or used, but one Pound Sterling in 1700 represents a different monetary system than the same unit of value two hundred years later. If a commodity is essentially free because of its abundance (as was the case with fuelwood in the rural areas of the United States during the last century), then that commodity also has no real economic price in the same sense as the air that we breathe. This of course does not mean that a cord of wood had no value at the point of collection. Wood was actually critical for survival but since it was in abundance, its price was insignificant representing basically the cost of timer cutting and processing.

In the following we will show a number of examples where the logistic substitution model can be applied because a common physical unit of measurement can be constructed. These examples do not cover a whole sector of the economy as in the case of energy substitution because of measurement difficulties. Instead we will focus on various methods of steel production and the tonnage of merchant fleets by various types of vessels.

2.4 Dynamics of Technology : Steel and Ships

The age of coal is usually associated with the development of three important leading sectors - iron and later steel industry, steam ships with iron hulls, and steam locomotives and railroads. The development of all three new technologies went hand in hand. The most voracious consumer of energy in the early periods of the industrial revolution was the blast furnace in which iron ore is reduced to pig iron. In Great Britain during the second half of the eighteenth century mineral coal was substituted for charcoal for smelting iron. This process was introduced in 1760 after the invention of making coke. A similar transformation of the iron industry in the United States was achieved one hundred years later during the middle of the nineteenth century. Figure 2.12 shows the substitution of fuels used in the smelting of pig iron in the United States. The first mineral fuel to be used in iron smelting was anthracite coal, which was introduced before bituminous coal in the United States because of easier accessibility and better transport possibilities. By the 1850s, one half of the energy used for iron smelting was provided by anthracite coal, the other half still being charcoal, but charcoal use was on the decline. During the second half of the nineteenth century anthracite was replaced by bituminous coal, which in turn was later substituted by oil and natural gas as the major energy sources in iron and steel production.

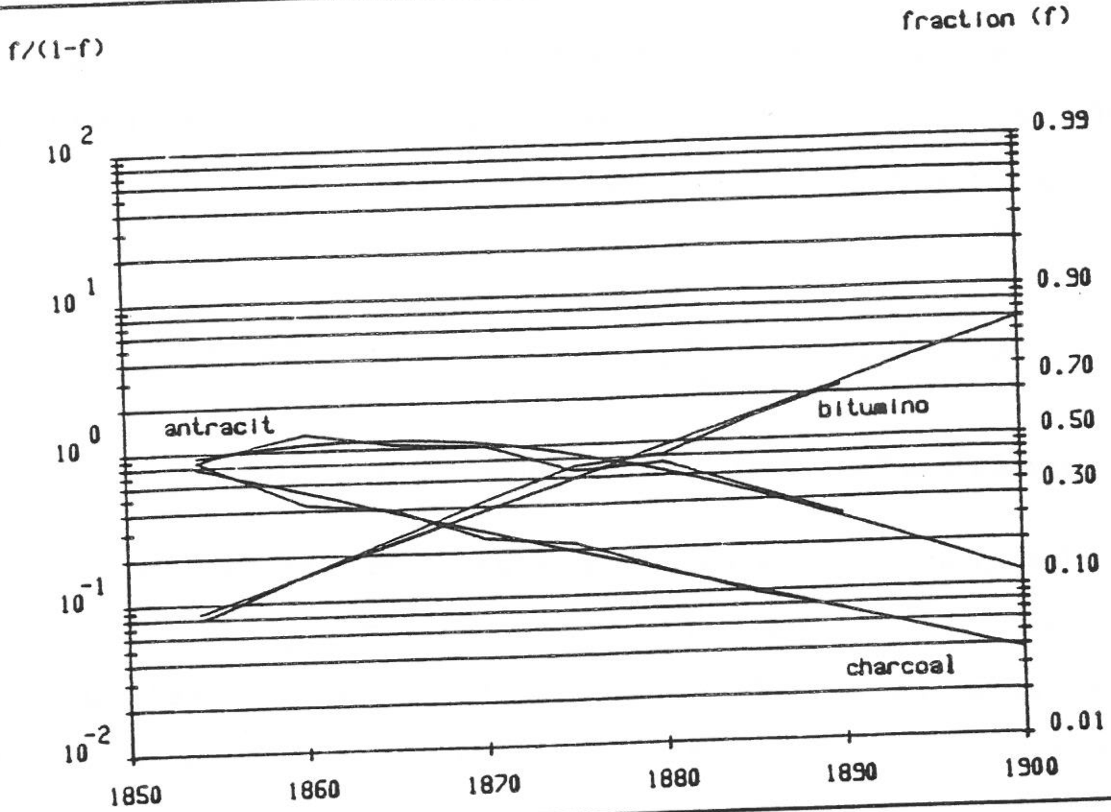


Figure 2.12 Fuel Substitution in Pig Iron Smelting, US.

The next important landmark in the history of the iron and steel industry was the introduction of the Bessemer process in the late 1860s for large-scale and low-cost steel production. By that time about ten percent of merchant ships were already powered by steam in the United Kingdom and the United States. The rapidly growing iron and steel industry made the substitution of metal for wood in building steam ships possible. The already extensive railroad network also provided efficient transport over continental distances that could mesh with faster and larger metal vessels. The new steel production process also provided superior and more durable steel for the rails. By the 1860s coal had also replaced wood as locomotive fuel in the United States. Thus, toward the end of the nineteenth century, both modes of long-distance transport, ships and railroads, were powered by coal and used steel as the basic construction material.

Figure 2.13 gives the raw steel production in the United Kingdom and Figure 2.14 in the United States, respectively. The figures show that in both countries production increased rapidly during the second half of the nineteenth century. The production increases were spectacular in the United States, so that America surpassed the United Kingdom during the 1890s in total production of both steel and pig iron, and ten years later also in the production of coal.

The rapid increase in the steel production became possible after Henry Bessemer patented the first high-tonnage process for steel production in 1857. This was an acid steel production process. Chemically, all steel production processes may be classified as acid or basic, depending on the refractory and slag combinations. The next improvement in steelmaking was achieved by the introduction of the open-hearth furnace. The first open-hearth to be used widely was based on acid chemistry although later the basic open-hearth also found extensive use. The basic systems have a decided advantage in flexibility with

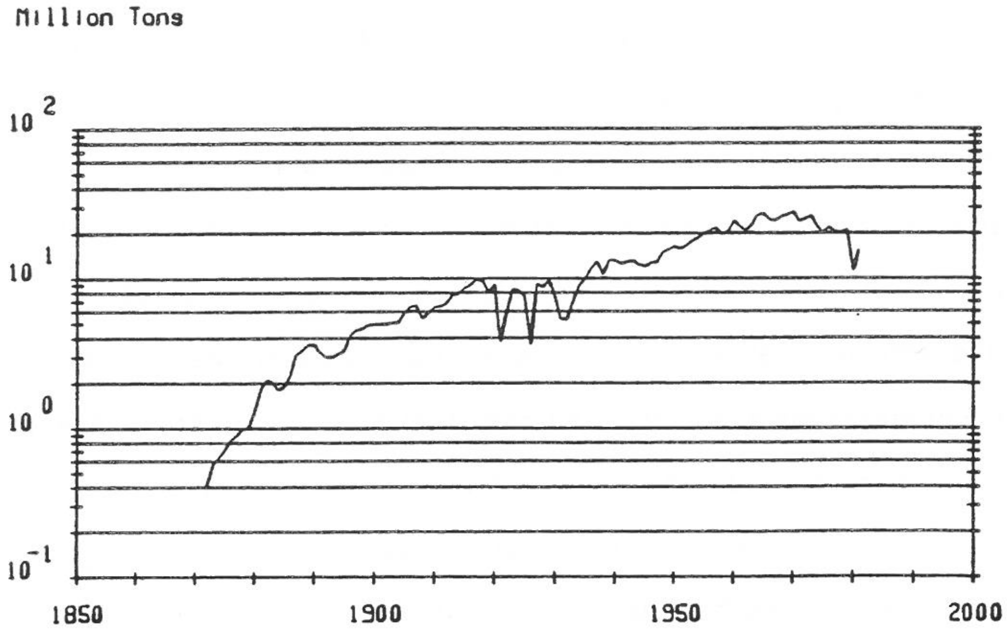


Figure 2.13 Raw Steel Production, UK.

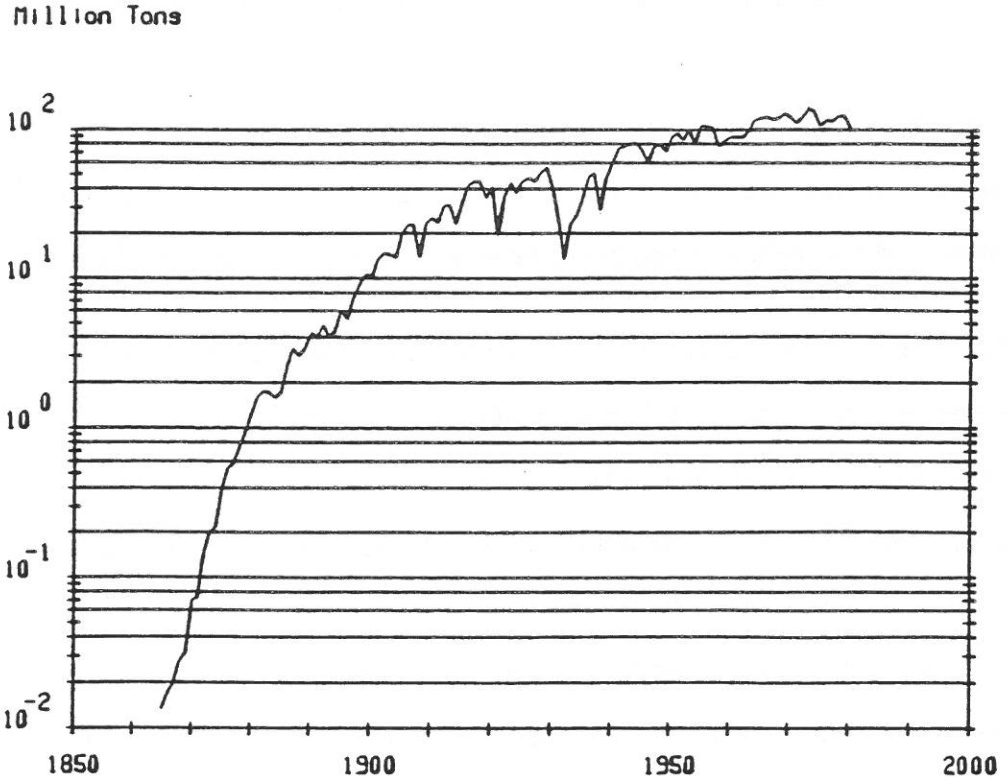


Figure 2.14 Raw Steel Production, US.

regard to raw materials consumed and grades of steel produced. The steelmaking processes were further improved by the use of oxygen for excess combustion instead of air. This offers many advantages such as faster melting and reduced checker chamber capacity. Consequently, the Bessemer process was also extended to basic chemistry and oxygen use, the most spectacular application originating in Austria as the Linnz and Donnewitz (L-D) process, now generally referred to as basic-oxygen steelmaking. The last improvement in steelmaking technology was the introduction of the electric arc furnace. The electric process has the advantage that it is suitable for making many grades of steel and can almost exclusively use recycled scrap iron and steel. Today it is used very extensively in the United States because the stagnating demand for steel allows for most of the production to be achieved by using scrap iron and steel (Miller, 1984).

Figure 2.15 shows the application of the logistic substitution model to the production of steel in the United Kingdom. The steel production has been disaggregated by the process-chemistry employed. Shortly after the introduction of the Bessemer process, acid steelmaking dominated production. During the 1920s, the rapidly growing use of basic processes already captured fifty percent of total output and the acid processes continued to decline reaching slightly more than a one-percent share in the 1940s. In the 1950s the basic processes provided almost ninety percent of all steel, but are now in decline after the introduction of the electric process. Since there is also a strong trend to favor the basic practice in electric steelmaking, the basic processes, taken together, provide almost all of the steel made in the United Kingdom today.

Figure 2.16 shows the actual technological substitution in steelmaking according to the process used in the United Kingdom and Figure 2.17 in the

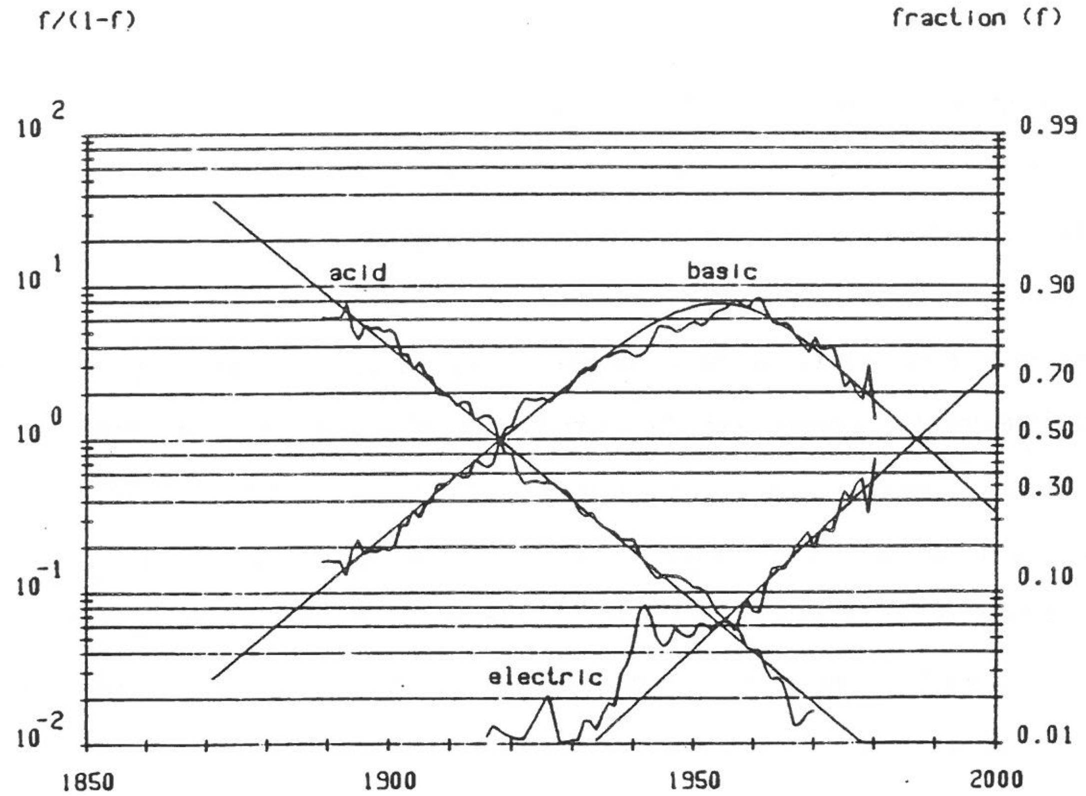


Figure 2.15 Substitution in Steel Production by Process-Chemistry, UK.

United States, respectively. Prior to the introduction of the Bessemer process all steel was produced by the traditional crucible methods used since antiquity. Data for the United States are available for the period before 1860, and Figure 2.17 shows that the Bessemer process replaced the traditional methods within two decades supplying almost ninety percent of all steel by 1880. From then on the Bessemer process was replaced in both countries by open-hearth steelmaking. In the United Kingdom the open-hearth process supplied fifty percent of all steel by 1885 and in the United States fifteen years later. The use of the open-hearth process continued to increase during the first decades of this century and by the 1940s accounted for more than ninety percent of the steel produced in the United Kingdom, reaching the same market shares about ten to fifteen years later in the United States. Thus, in both countries the technological

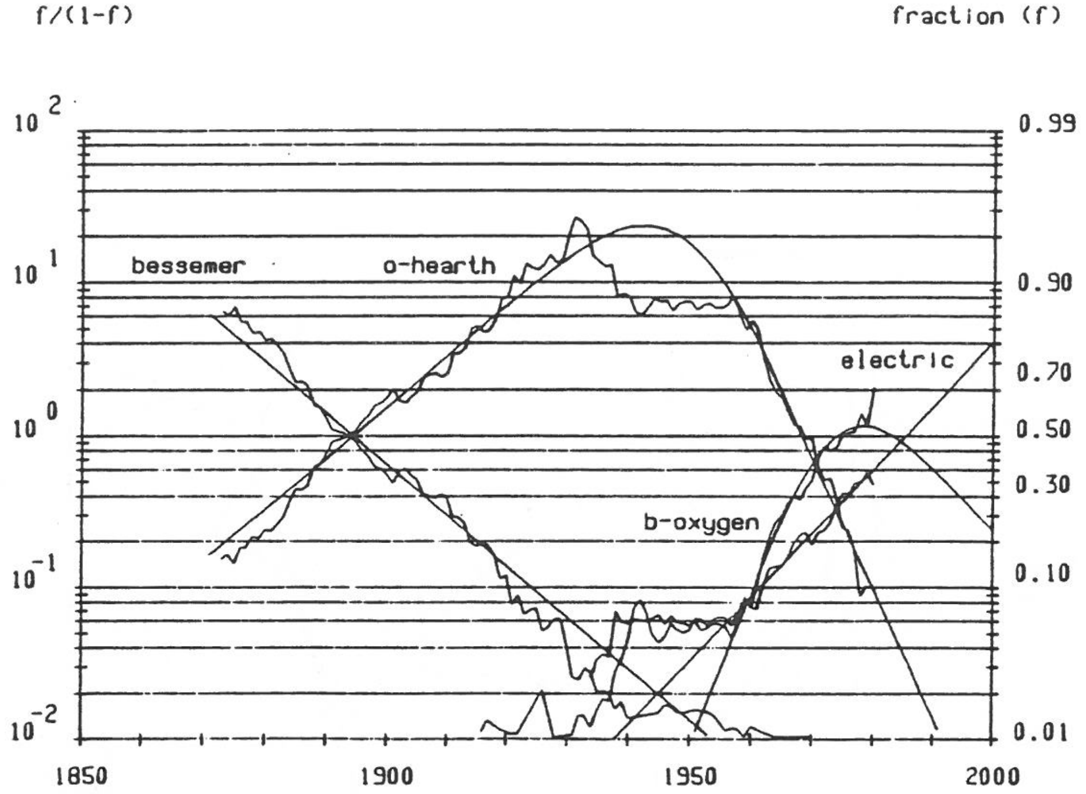


Figure 2.16 Technological Substitution in Steel Production, UK.

substitution of steel production shows similar development. This parallel development has continued during the last forty years with the introduction of the basic-oxygen and electric steel processes. The electric arc process was introduced as early as 1900, so that it gained importance before the basic-oxygen process. In both countries, however, the basic-oxygen process expanded faster, probably because it is technologically similar to the open-hearth and Bessemer basic variants. During the 1960s, the basic-oxygen process portrayed very rapid share increases reaching more than fifty-percent of the market in the 1970s. Accordingly, once the most important steelmaking process, open-hearth declined rapidly down to the ten-percent mark during the same period. The electric process is gaining importance and will probably overtake basic-oxygen within the next two decades. Due to the saturation of demand for domestic steel in the

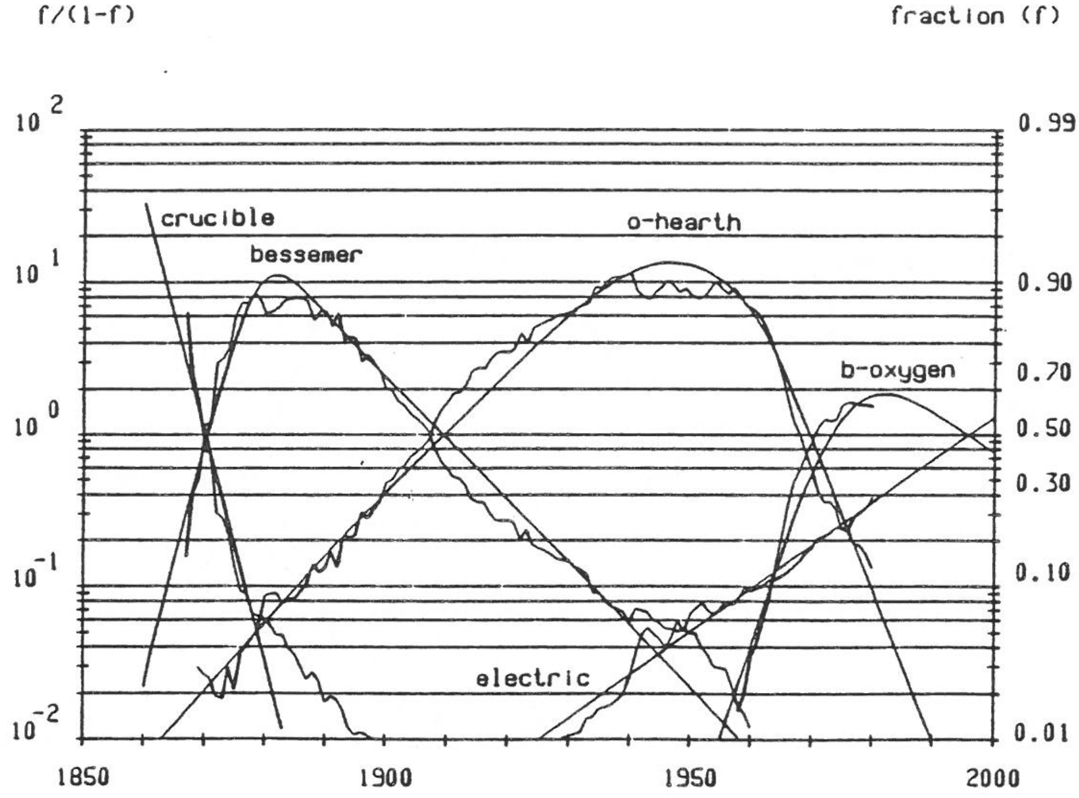


Figure 2.17 Technological Substitution in Steel Production, US.

United Kingdom and the United States it is conceivable that the electric process may achieve the importance that the open-hearth and Bessemer once had. The dwindling total production in these two countries leads to higher and higher percentages of scrap iron and steel inputs instead of iron ore in addition to some imports of pig iron. This development favors the electric process since it is very energy efficient and allows for almost exclusive use of recycled inputs (see, Miller, 1984).

These examples illustrate that the evolution of the steelmaking technologies portray a regular pattern that is similar to energy substitution. The decomposition of production in the United Kingdom according to the chemistry of the steelmaking process used, shows that technological substitution is invariant to the different characteristics of the market under analysis. We have also shown that the energy inputs to steel production in the United States follow logistic substitution trends. The description of the historical data by the logistic substitution model was consistently accurate, in spite of definitional problems encountered in the decomposition of steel production by different technologies. Especially the departure of the actual substitution trends from the estimates in the United Kingdom during the 1930s and the 1940s, in Figure 2.16, indicate a possible accounting error. It is possible that a part of the steel production allocated to the basic-oxygen process during this period was actually a part of the basic open-hearth technology. Only careful analysis of more detailed data could indicated whether this *ex post* "model prediction" could be validated. This hypothesis is strengthened by the fact that a similar departure from the long term trend is not present in the substitution dynamics when expressed in terms of chemical processes in Figure 2.15.

The evolution of the merchant fleet in the United Kingdom and the United States can be analyzed with respect to two related technological substitution processes. The first refers to the type of propulsion used and the second to the construction materials. The traditional propulsion used ever since ancient times was wind power and the traditional construction material was wood. With the development of the steam engine and the relatively high energy density of high-quality coals, it was possible to slowly replace sails with steam engines. The first designs were of a hybrid type employing both steam and wind power. With the increase in the size of vessels that was necessary because of the expansion of overseas trade, and with the growth of the iron and steel industries, wood was increasingly substituted by iron and later steel as the basic construction material. In fact, the number of vessels remained practically constant since the end of eighteenth century at about 25 thousand ships in both countries and only in the United States has it doubled during the last three decades. During the same period of almost two centuries the total registered tonnage of merchant fleets increased by almost two orders of magnitude (in the United States), implying that the average vessel is about hundred times larger today than in 1800. This enormous increase in the tonnage capacity of an average vessel can only be explained by continuous improvements in propulsion systems, construction materials and design.

Figure 2.18 shows the tonnage growth of the merchant fleet in the United Kingdom since 1788 and Figure 2.19 in the United States since 1789. Unfortunately, the data do not extend beyond 1938 for the United Kingdom, because ships of less than 500 tons were not accounted for any more. Figure 2.20 and Figure 2.21 show the substitution of sailing by steam ships, both coal and oil fired, and later the market penetration of motor, diesel and semi-diesel ships in the United Kingdom and the United States, respectively. Sailing ships dominated the merchant fleets in both countries until the 1830s. Steamers acquired a one-percent share of the total tonnage in 1819 in the United States and in 1826 in the United Kingdom. In spite of the later start in the United Kingdom, sailing vessels were replaced somewhat faster than in the United States, so that by 1880 steam ships accounted for one-half of the merchant tonnage whereas in the United States the same share was achieved ten years later. By the 1920s steamers constituted more than ninety percent of merchant tonnage in both countries. During the same period motor ships were also introduced. Their shares of total tonnage has increased in both countries ever since, although even today they have not acquired much more than one tenth of the total tonnage in the United States. Thus, steam ships remain an important type of merchant vessel, although today they are fueled by oil and in some cases use steam turbines instead of coal fired atmospheric engines. Nevertheless, their share is still decreasing in favor of motor ships. The substitution process in the tonnage of merchant fleets was

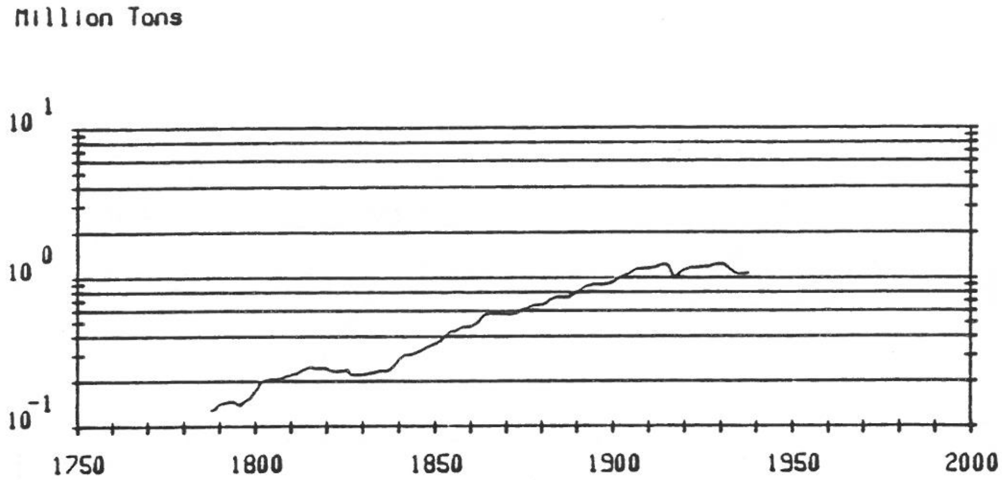


Figure 2.18 Tonnage of Merchant Vessels, UK.

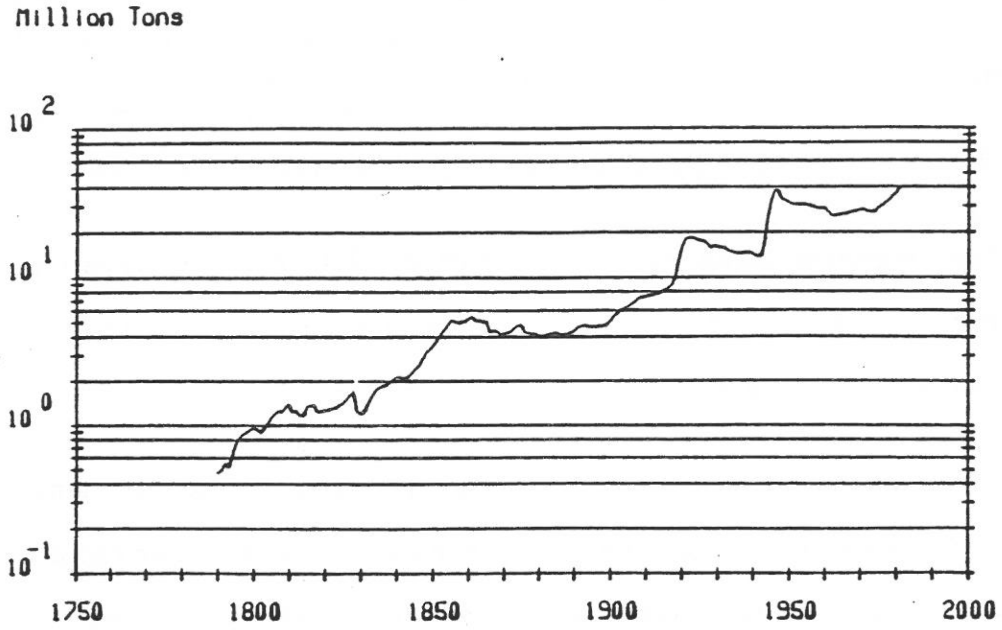


Figure 2.19 Tonnage of Merchant Vessels, US.

somewhat more dynamic in the United Kingdom than in the United States. The replacement was also more stable with respect to the logistic substitution paths derived from the model. During the Second World War, the share of motor ships sharply increased in the United States and accordingly the share of steamers was

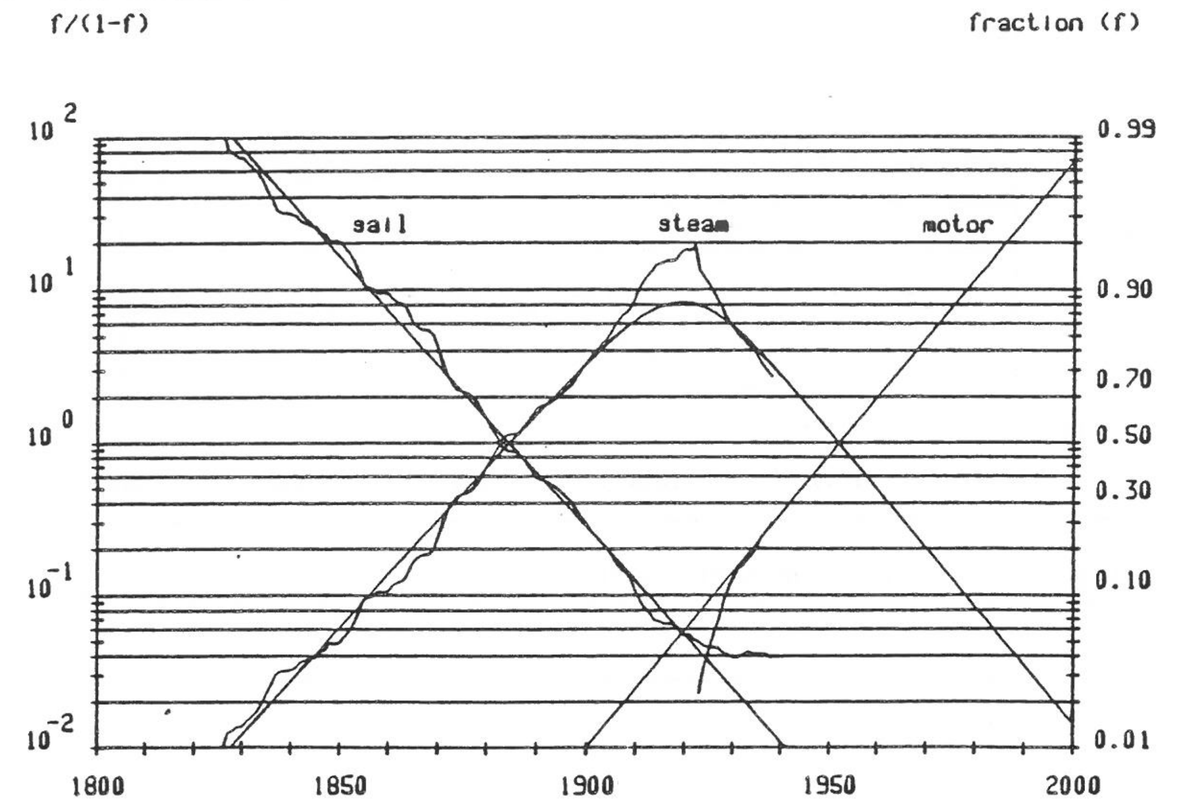


Figure 2.20 Substitution in Merchant Vessels by Propulsion System, UK.

below the long-term trend during this period. But these perturbations were reabsorbed during the 1960s to return to the long-term trend indicated by the logistic substitution model.

Figure 2.22 shows the shares of wood and metal ships in the merchant fleet of the United States. The replacement of wooden ships was a rather rapid process that started soon after the introduction of Bessemer steelmaking. The first metal ships were made out of iron but later also steel was used. The data are not available for the period before 1885 so that we can only extrapolate, using the logistic substitution model, that metal ships achieved the one-percent share around the year 1850. Already by 1910 fifty percent of all merchant tonnage consisted of metal ships and today virtually all ships are made out of metal.

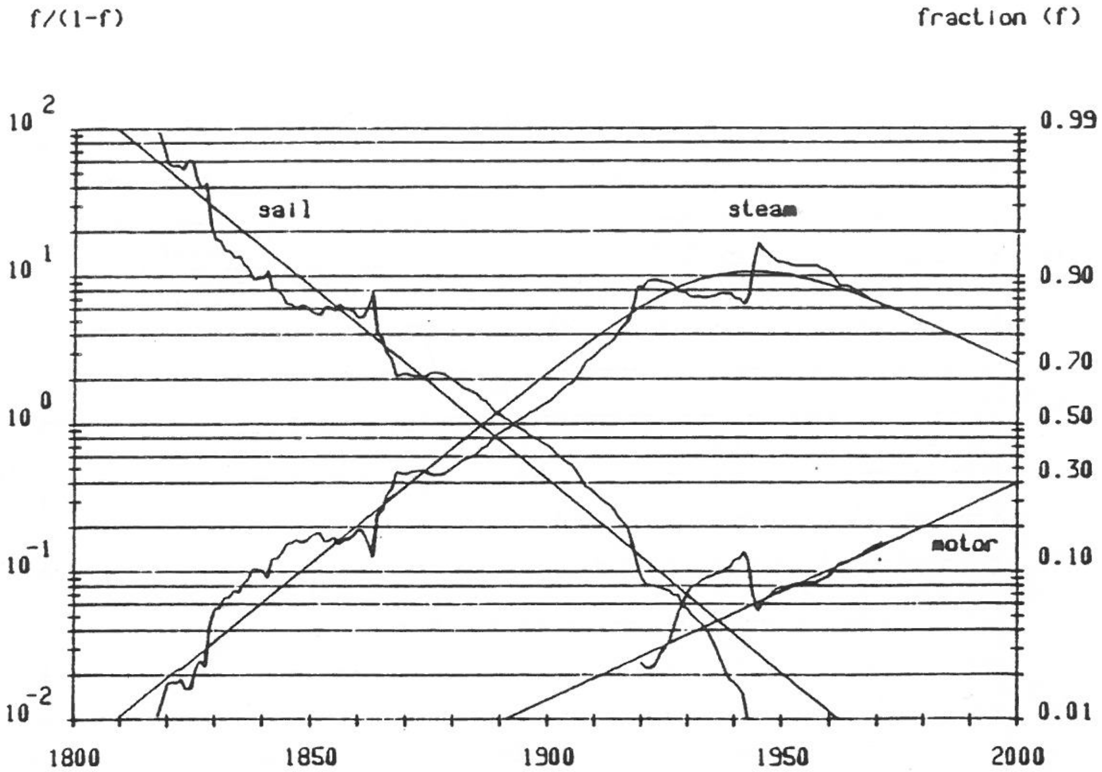


Figure 2.21 Substitution in Merchant Vessels by Propulsion System, US.

The above applications of the logistic substitution model to the historical replacement of new for older forms of energy, methods of steel making and propulsion of merchant vessels have all indicated that improvements and growth are achieved through a regular but not continuous process. From the time of its first commercial use each new improvement or technology grows logistically until it reaches a saturation point and then proceeds to decline logistically while being replaced by newer and more promising technology. During each phase of the substitution process the dominant technology appears to be strong and unassailable, but with time it decays as emerging competitors "attack" the newly exposed position of the mature technologies. It is interesting to note that the saturation point is not by and large determined by mere physical or resource limitations but rather through the dynamics of the introduction of new

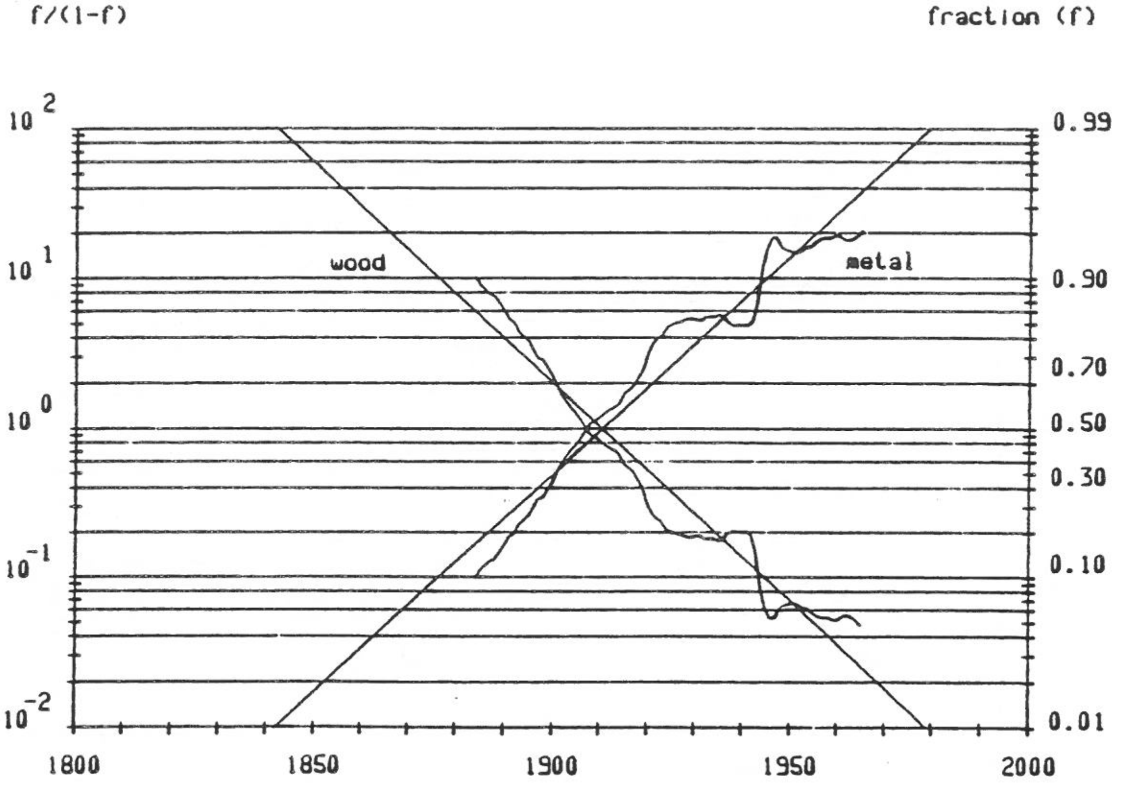


Figure 2.22 Substitution in Merchant Vessels by Structural Material, US.

technologies. Thus, the market shares increase until limits are encountered that appear to be endogenous to the market itself. These limits are encountered before complete market takeover and before resource depletion. Although it became scarce in many areas fuelwood was not exhausted before it was replaced by coal. In the United States the supply of wood was practically inexhaustible. Thus, the introduction of coal was to some degree necessary for other reasons that are independent of fuelwood depletion. If we apply this argument to the British experience we could conclude that fuelwood was not exhausted in the classical sense of the word in Great Britain either. Instead it became more difficult to provide sufficient amounts of wood that would have been required by the rapid growth of production and energy demand. In principle, wood could have been imported from more distant locations, but this would have been very costly due to long, cumbersome transport requirements. The sharp rise in fuelwood prices in Great Britain also indicated the difficulties of supplying wood to match the increasing energy demand. Due to its higher energy density per unit volume, coal was easier to transport and was thus introduced as the major energy source. Coal also allowed more concentrated production because it offered the possibility of removing the production sites from energy extraction areas. The advantages of easier-to-transport energy sources with high energy densities (per unit weight or volume) became more obvious when crude oil replaced coal. Oil is transported over global distances at very low cost and with its emergence it was possible to further concentrate production and make it practically independent of the energy source location.

Each newer energy technology also offered the possibility of improving the efficiency of energy use. The coal-based energy system was much more efficient than those fired by wood. Coal-fired locomotives are more efficient and can be made more powerful than the wood-fired ones. Similar efficiency improvements were possible through increased use of oil and later natural gas. In power generation these improvements are very clearly observable. Coal power plants are less efficient than oil-fired ones, and natural gas turbines offer the most efficient process for electricity generation. Diesel and electric locomotives are much more efficient than steam ones. Thus, the necessity for more concentrated, easier to transport, versatile and efficient forms of energy were primary reasons for the substitution of older by newer forms of energy. With the further increases in production and overall economic activity all of these improved qualities of the newer energy forms became necessary. We argue here that technological improvements and the expansion of new technologies was not a random process driven by chance, but rather a necessity caused by economic and population growth. The use of each energy source grew to limits imposed by the growth process itself. After a certain point it was no longer feasible to simply increase the energy inputs and flows, thus new sources had to be introduced that were more manageable at increased levels of use. The restructuring of the whole economy caused changes in energy supply. The further growth of older energy forms encountered limits imposed by the dynamics of economic development and caused their substitution by newer sources that could fulfill the new demands more adequately.

This argument can be generalized to other substitution processes that we have analyzed. The introduction of new steelmaking processes and improvements of the existing ones were necessary because of the growing use of steel. Substitution, powered by increasing use, is a kind of autocatalytic process. Growing demand makes the introduction of a new process desirable, but once it is introduced it also allows new demand increases that would have not been possible with older methods. Cruicible steel making was too expensive and could not provide for large-volume production. The Bessemer process found wide application immediately after it was patented because the demand for steel was very high. New basic processes allowed more flexible output mix and with demand decreases electric methods have provided for the economic use of abundant scrap steels. Thus, the application of the electric method illustrates a technological change that is induced by the structural change in the market and growth pattern.

Finally, the technological improvements of propulsion and structural materials of merchant ships illustrates a similar evolution. Since the number of ships remained practically constant, the increased volume of transoceanic trade necessitated increases in the size and speed of vessels. Increased speed was possible only through more power since the vessels also increased in size. Thus, the engines had to be made more efficient and the energy form had to have a higher density in order to minimize storage problems. A larger size also called for stronger and easier to manufacture materials. Given these constraints it was obvious that sailing vessels made out of wood could not survive in spite of the fact that the limits in wind "supply" were not reached and that wood for shipbuilding was still abundant in many parts of the world. Instead, the use of traditional sailing vessels reached limits imposed by new demands on marine transport. With the availability of new construction material and propulsion systems the older ones were replaced. It is still somewhat mysterious why the number of vessels remained practically constant over such long time periods. Probably it is due to the fact that larger and faster ships could better accommodate the increasing volume of marine transport while decreasing transport time. This could be linked to the question of logistics - had the average size of vessels remained constant over the two centuries, the number of ships would have had to increase ten times because the tonnage of the fleets increased by about an order of magnitude. With so many ships around, safety, scheduling, loading and unloading, manpower and other aspects of marine transport could have become unmanageable.

The technological change in energy, industrial production and transport also appears to be interlaced and related. With hindsight they necessitated each other since the expansion of a given new technology also required the expansion of many other new technologies. Transport systems had to improve to allow for higher trade and production. Higher trade and production made more efficient transport possible through use of new materials, energy sources and so on. It is obvious that this kind of reasoning leads to a circular argument that everything depends on everything else. Nevertheless, the technological substitution dynamics illustrated that the changes have a regular pattern and rules that point to a certain rhythm in the structural change of human activities. Every growth process continues to limits imposed by the structure of a given market that is in turn related to overall economic and social development and not necessarily to mere resource depletion. Once these limits are reached further growth becomes economically and socially unviable. Horse riding, wood fire and sailing ships have become aesthetic and recreational activities in the developed economies while they still constitute a daily necessity in many developing parts of the world as means of transportation and source of energy.

3 THE LONG WAVE IN ECONOMIC DEVELOPMENT

We have seen that technological advancement is an evolutionary but not continuous process. Innovations, and their dissemination until they became common practice, came in waves of substitutions with a duration in the order of fifty years for large systems and infrastructures. It is therefore only natural to ask whether the whole process of economic growth and development also portrays equivalent behavior. The question is whether economic development can also be considered as a series of leaps with periods of rapid growth and periods of relative stagnation. From history we know that this is at least an approximate description since a number of serious depressions and crises as well as periods of unusual prosperity and great achievements have been recorded since the beginning of the industrial revolution.

3.1 Early Research

The identification of the fluctuations in economic growth, longer than the business cycle, is generally credited to Kondratieff (1925 and 1928), often transliterated from Russian as Kondratiev. The discovery of the long wave in economic development dates back to Socialist economists in Germany and the Netherlands, notably Parvus (1901 and 1908), van Gelderen (1913), who wrote under the pen-name Fedder, and de Wolf (1924 and 1929). While Kondratieff acknowledged two of his predecessors he was the first to use empirical evidence in an attempt to confirm the existence of the long wave. He analyzed 25 historical time series, some of them indicating rising and declining movements with a period between 48 and 60 years. Kondratieff was successful in showing strong evidence of the long wave in price movements, but was criticized for making a premature conclusion that they are also evident in "real" indicators. His method consisted of first eliminating the secular trend from the time series. The trend was eliminated directly by the method of least squares from the price and interest rate series. The physical and quantity indicators were divided by the population and then the trend was eliminated by the same method. The deviations of the actual data from the estimated trend were then smoothed by a nine-year moving average in order to eliminate the business cycle and other fluctuations shorter than the long wave. The smoothed deviations from the trend indicated in many series the evidence of the wave-like fluctuations.

Although Kondratieff stated his hypothesis very cautiously, namely that "we consider the long cycles with an average duration of about fifty years in the capitalist economy only as probable" (Kondratieff, 1926), he encountered strong criticisms in his own country in 1922 after the publication of his first paper on long cycles. Garvy (1943) argues that this strong criticism helped Kondratieff to refine the empirical foundations of his work. Kondratieff also offered a number of possible explanations for the "long cycles in the economic life" emphasizing the discontinuity in the production and duration of long-term fixed capital, but he did not provide an appropriate theory of long waves.

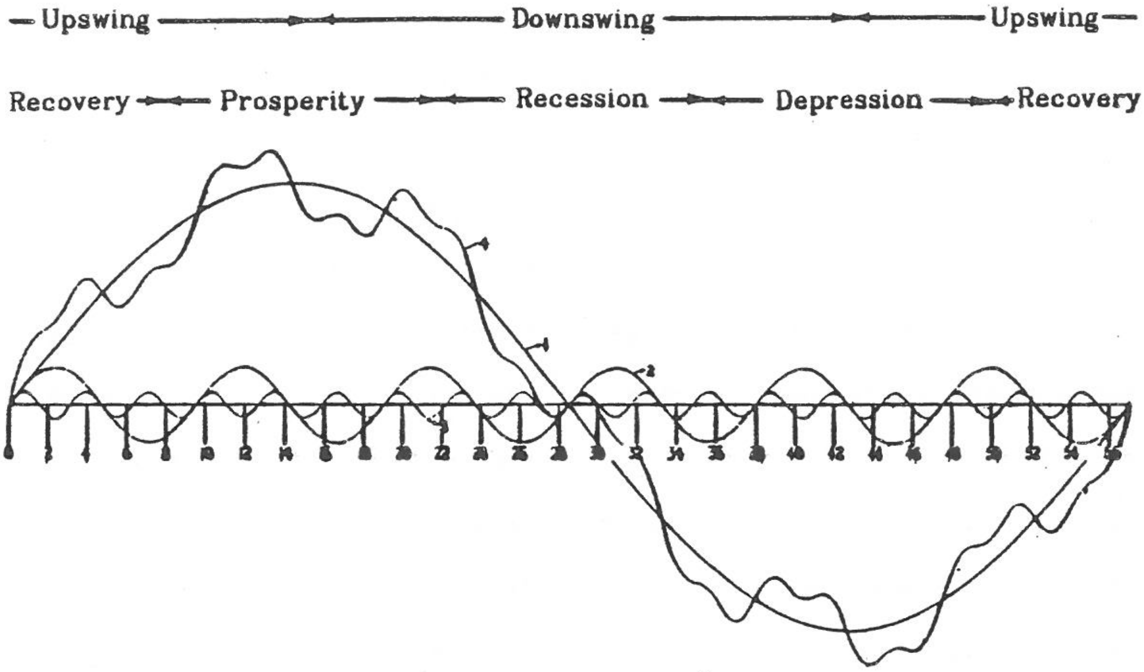
The long wave in economic development has remained to be a controversial issue to date, both with respect to the empirical evidence of its existence and with respect to possible explanations and theoretical foundations. In fact, most of the subsequent research on long waves concentrated on possible theoretical formulations of the long wave mechanism rather than on the empirical evidence of its existence. Van Duijn (1983) noted that this dispute is not surprising because the duration of the long wave itself makes the empirical validation very difficult. Statistical series of at least one if not of two hundred year duration must be used if more than two waves are to be observed. The available statistical tools are not suited for the analysis of such long fluctuations in the time series.

In recent years, there has been a renewed interest in long waves and their theoretical explanations probably due to the slowdown in economic growth during the last decade. During the first three decades of the postwar period, the "apparent evenness and durability of economic growth" made "any explanation of growth from a long-wave perspective seem absurd" (Van Duijn, 1983). Consequently, the research on long waves can be divided into two periods, one following the pioneering work of Kondratieff and his predecessors and the second starting a decade or so ago.

Schumpeter was the most prominent proponent of the long wave hypothesis. In his famous and long-forgotten book on "Business Cycles" Schumpeter (1939) described the process of economic growth and technological progress as a succession of long waves each initiated by expansion of leading sectors that emerge from innovating enterprises under the competitive business environment. Schumpeter considered the clustering of innovations during the period of emergence of new leading sectors as the explanation for the wave-like form of economic development. For Schumpeter, innovations come in clusters, and are not evenly distributed or continuously absorbed, due to the basic principles that govern the process of capitalist development. "As soon as the various kinds of social resistance to something that is fundamentally new and untried have been overcome, it is much easier not only to do the same thing again but also to do similar things in different directions, so that a first success will always produce a cluster" (Schumpeter, 1935). Consequently, entrepreneurial activity is also clustered "because the appearance of one or a few entrepreneurs facilitates the appearance of others, and these the appearance of more, in ever increasing numbers" (Schumpeter, 1934). Schumpeter's definition of innovation encompasses "changes in technique of production, the conquest of new markets, the insertion of new commodities, and so on. This historic and irreversible change in the way of doing things we call innovation and we define: innovations are changes in production functions which cannot be decomposed into infinitesimal steps. Add as many mail-coaches as you please, you will never get a railroad by so doing" (Schumpeter, 1935).

This concept of clustering of innovations and the emergence of leading sectors is very closely related to the market substitution processes that we have analyzed. Each new technology is based on a host of innovations and necessitates the development of new infrastructures once it becomes successful in replacing the traditional technologies. The success of one substitution process stimulates substitutions in other markets, so that they are all interrelated at the level of the whole economy. This swarm-like appearance of new technologies and their integration in new sectors "easily and necessarily explains the fundamental features of periods of boom. It explains why increasing capital investment is the very first symptom of the coming boom, why industries producing means of production are the first to show supernormal stimulation" (Van Duijn, 1983). This growth necessarily leads to limits and eventual decline. The seeds of decline originate in the growth phase of the wave. The emergence of new competitors and their demand for means of production tends to increase the prices. Thus, costs rise with the onset of prosperity, especially in the old enterprises that operate with traditional means of production. As a result earnings are slowly reduced, first in the old enterprises but later also in the innovative ones. The timing of this process then also determines the length of the boom. It essentially reduces to the time it takes for innovative methods, if they are successful, to acquire substantial shares of the market and possibly also create new markets. Once the innovators have captured the markets, prices tend to fall, demands are saturated. These events mark the end of the prosperity period. Stagnation and later depression are further stimulated by the decrease in borrowing caused by the repayment of debts by successful innovators during the prosperity phase. This in turn reduces the purchasing power and reduces the demand at the time when productive capacities are largest. As profits decline, monopolistic markets slowly disappear, competition intensifies and induces rationalization in order to reduce costs. The liquidation of the excess productive capacity depresses the prices further and leads again toward a drive to innovate. Stagnation leads to depression and may lead to a crisis. In a strive to improve their competitiveness, individual entrepreneurs start adopting innovations at the beginning of the depression period. This innovation process eventually produces another long wave.

The Schumpeterian chronology of the long wave consists of four distinct phases - prosperity, recession, depression and recovery. Figure 3.1 illustrates these four phases of the long wave together with the other two shorter cycles. In fact, Schumpeter considered the economic evolution to consist of complex contours of superimposed cyclical movements. For the sake of simplicity, he explicitly considered three overlapping cycles - the Kondratieff cycle with a duration of about fifty years; the Juglar, business or investment cycle with a length of seven to eleven years; and the Kitchin or inventory cycle with a length of three to five years. Consequently, he disregarded all the other types of fluctuations "as an illustration of all the boldest assumptions which it is possible, and to some extent permissible, to make in order to simplify description and to construct an ideal schema with which to compare observations" (Schumpeter, 1939). He recognized that the three-cycle schema did not imply that the effects



Curve 1, Long Wave; Curve 2, intermediate cycle (Juglar); Curve 3, short cycle (Kitchin); Curve 4, sum of 1 to 3.

Figure 3.1 Schumpeterian Model of Long-Wave Phenomenon (Schumpeter, 1935).

of the fluctuations are simply additive. He was also cautious to note that no rational justification is evident for assuming that the integral number of Kitchins in a Juglar or of Juglars in a Kondratieff should always be the same. Yet, he was convinced that the historical time series indicate that there are six Juglars in a Kondratieff wave and three Kitchins to a Juglar cycle. It is evident that even this simple schema of economic development, disregarding other fluctuations, would lead to a very complex and irregular-looking composite pattern of development.

Finally, we should also observe that Schumpeter considered the fluctuations in economic development to constitute departures from the equilibrium. During each cycle the system would strive to achieve an equilibrium, but this ideal state of the system is never reached due to new changes in the system. The process of economic development consists of changes that are not exogenous, but rather arise within the economic system itself. The most critical of these changes can be seen in the innovation process that leads to the creation of new leading sectors and opening of new markets. Thus, the clustering of innovations and innovative entrepreneurial practice has a powerful impact on the pattern of economic development, first it leads to "gales of creative destruction" and later to renewal and growth. Another way of understanding this dialectic succession of alternate phases of development is that the economic process oscillates between a tendency toward an equilibrium during the upswing phase and a tendency to move away from achieving this equilibrium state during the downswing. Andersson (1984) called this aspect of economic process the equilibrium attraction and repulsion: When the economy is close to a potential equilibrium, during the prosperity phase, profits tend to fall causing the entrepreneurs to strive for new profit opportunities. New profit opportunities appear in the form of innovative practices which disturb the process of achieving an equilibrium state and initiate a renewed process of growth toward a new equilibrium after a period of stagnation. Thus, the economy pulses - first it strives toward an equilibrium (attraction), but when it gets close to achieving this abstract state, the economic system undergoes structural changes that are initiated by the activities of the entrepreneurs (repulsion) and strives toward a new state of equilibrium. This explains the clustering of innovations and emergence of new enterprises during the downswing period. If the innovations and new enterprises were to appear independently of each other, they would probably be evenly distributed over time. In such a case "there would be no boom and no depression as special, distinguishable, striking, regularly recurring phenomena" (Schumpeter, 1934). Since innovations tend to cluster, the changes resulting from innovations are usually not small through the additive and interrelated effect of new entrepreneurial practice. Consequently, "the new does not grow out of the old but appears alongside of it and eliminates it competitively", changing the general economic conditions so that a special process of adaptation becomes necessary (Schumpeter, 1934).

This description of the process of economic change parallels our analysis of technological substitution. We have seen that new energy sources, steelmaking methods and merchant vessels represented genuine innovations that did not evolve out of the old practice, but rather competed with the old in a remarkably regular pattern. Thus, the Schumpeterian explanation of the long waves in the economic development and the dynamics of logistic substitution are inherently related and represent complementary aspects of the same process.

3.2 Toward a Theory of the Long Wave

The apparent evenness and the continuity of steady economic growth after the Second World War probably caused economic theory to emphasize long-run growth primarily under steady state and equilibrium conditions. Most of the research efforts were directed toward fine-tuning the ever-expanding economies. The neoclassical theories were perfected and reached high levels of formal rigor and internal consistency. This no doubt caused their powerful impact on economic thought in general and their widespread popularity and application in particular. Consequently, many earlier analyses of the dynamics of economic development, including the long waves research, have fallen into disuse and receded to the margin of economic thought. Particularly numerous and varied are the theories of the pre-war period that have attempted to identify cycles in the growth process. Much of the work concentrated on empirical analysis of the dynamics of economic growth, prosperity and recession. It suffices here to mention among the many researchers Mitchell (1927), Burns (1934), Kuznets (1930) and Hoffmann (1940). Many theories were developed to explain the discontinuous nature of economic growth, some emphasized monetary aspects and gold supply, others the role of saving and investment including the shortage-of-capital theories, the influence of innovations and emergence of leading sectors, wars and so forth. Especially the work of Kuznets toward a general, dynamic theory of primary production and prices is well known and has led to the formulation of the Kuznets or building cycle with a duration of 20 to 25 years. Kuznets found that production and price trends reflected systematically the life-cycle of important technical innovations or new markets in sectors that had the "lead in development". A cycle consists of two phases. During the first phase, production increases rapidly and price decreases rapidly. In the next phase, the rate of change decelerates diminishing price decreases and production increases. Thus, the cost-reducing effect of innovations is generally translated into price reductions, but over time diminishing returns prevent further price reductions and expansion of production.

During the last decade a renewed interest in the dynamics of economic growth and long waves emerged as world economies entered a period of prolonged recession. As interest in long waves has risen, a number of leading researchers have expressed doubts concerning the adequacy of the neoclassical theory to deal with the discontinuous growth associated with recession and structural changes experienced throughout the economy. For example, Tobin and Nordhaus (1972) have questioned the relevance of the neoclassical theory because it "conceals, either in aggregation or in the abstract generality of multi-sector models, all the drama of the events - the rise and fall of products, technologies, and industries, and the accompanying transformation of the spatial and occupational distributions of the population. Many economists agree with the broad outline of Schumpeter's vision of capitalist development, which is a far cry from growth models made nowadays in either Cambridge, Mass. or Cambridge, England" (Tobin and Nordhaus, 1972). It is fair to say that the prolonged recession of the last ten years caused a new awareness that "certainly revolutionized economic thinking, shattering, in various respects, the dearly beloved paradigm of equilibrium growth" (Bruckmann, 1983).

As the result of such doubts about the adequacy of traditional economic theory to describe the changing nature of economic development from prosperity to recession, many of the almost-forgotten long wave theories have been revived and new approaches proposed. It is probably too early to speak of different schools in the long wave research, but for the time being many of the fundamental issues are not resolved. A recent international conference on long waves held in Siena/Florence (see, Bianchi, Bruckmann and Vasko, 1983) indicated that the current controversy focuses on the questions of measurement and empirical evidence of the long-wave phenomenon, and on clarification of possible causes and theoretical explanations of the long fluctuations in economic development. Delbeke (1981) devised an interesting classification of the proposed theories according to the postulated, basic causes of the long wave phenomenon: innovation, entrepreneurship, capital, labor and raw materials. This classification scheme is useful as a guideline, although recently the emphasis has shifted toward multicausality as a more adequate approach to the complex phenomenon of the long wave (Bruckmann, 1983). Van Duijn (1983) analyzed both the divergent and complementary aspects of the long wave research efforts, and earlier Rostow (1975) made an attempt to synthesize several different approaches. Delbeke also concluded that his broad classification actually reveals more complementarity among proposed theories than the chosen headings would indicate. In fact, the list of only a few major headings for possible causes of the long wave already indicates some consensus among the researchers. We will briefly outline this classification scheme as presented by Delbeke (1981) and Bruckmann (1983) with some minor modifications.

3.2.1 Innovation and Entrepreneurship

Many long-wave researchers have undertaken a more rigorous reformulation of the Schumpeterian theory of economic growth by treating technological and entrepreneurial innovations as the primary driving force of economic development. Among others Mensch (1975), Marchetti (1983, 1984), Kleinknecht (1983, 1984), and Bieshaar and Kleinknecht (1984) emphasize the role of clusters of technological innovations in generating new growth sectors and markets. Mensch distinguishes between basic and improvement innovations. The basic innovations are essential for the creation of new leading sectors during the depression and they cause the creative destruction that eventually leads to prosperity. These basic innovations become accepted because entrepreneurs seek new investment possibilities, even if they may be risky and uncertain, due to demand saturation for products of old leading sectors. Thus, the acceptance of basic innovations generates rapid economic development of the newly established sectors and recovery is initiated. From this point on the usual life-cycle of industrial development begins in the leading sectors. This growth pattern resembles the life-cycle of leading sectors formulated by Kuznets with the initial rapid pace of expansion and eventual decrease due to diminishing returns. The growth path can be described by the S-shaped curve with basic innovations occurring before the initial rapid production increases. As the sector expands, further technological development is necessary and according to Mensch it is realized through improvement innovations. As the expansion diminishes toward the end of the life-cycle due to increasing demand saturation, a concentration and rationalization of the sector begins. The emerging monopolies protect their market niches through pseudo-innovations causing further price increases. A corner stone of the theory is that in this saturation phase pseudo-innovations become more important than the basic innovations and a phase of "stalemate in technology" is initiated. Thus, a long wave is completed with resulting recession. Increasing market saturation forces the entrepreneurs to invest less in production and more in money and capital markets. The recovery occurs only after new basic innovations offer new investment opportunities, and the resulting emergence of new sectors generates new demand. Thus, according to Mensch the only cure for stagnation and prolonged depression is the support of research that results in basic innovations because during the critical period of stalemate in technology the resumption of economic development depends on the creation of basic innovations and new sectors.

Mensch determined empirically the four major clusters of basic innovations, probably constituting one of the most important contributions toward understanding the long wave phenomenon. Figure 3.2 shows that these four major clusters of basic innovations occurred around the years 1775, 1825, 1896 and 1935 coinciding closely with four severe depressions. Most of the basic innovations that constitute the clusters were instrumental in developing the new sectors that assumed the leading roles in the subsequent phases of rapid economic growth. One of the difficulties with this approach is that it is not always obvious how to distinguish between basic and improvement innovations. Another criticism of the empirical analysis was advanced by Clark, Freeman and Soete (1981) who claim that the innovation data used by Mensch do not represent a

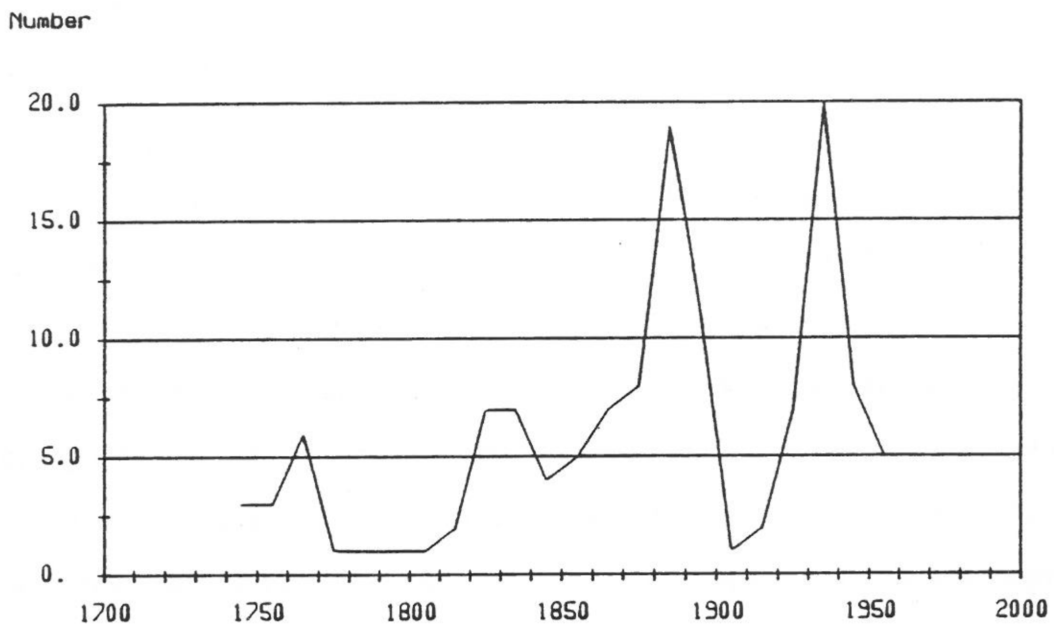


Figure 3.2 Frequency of Basic Innovations by Decade (Mensch, 1975).

reliable sample. On the other hand, even if the data set is biased, it was collected for different purposes so that there is no *a priori* reason to assume that this particular set is inadequate as evidence for the clustering hypothesis. A further difficulty is that basic innovations can only be identified *a posteriori* once they have helped the creation of a new sector. Many innovations, although they may appear to be of fundamental importance, never lead to the creation of new sectors. Some innovations may be revolutionary in character, but if they occur during the growth phase of the wave, they would only lead to some improvements in the already mature sector. Thus, the depression offers the opportunity for creating new sectors with the timely implementation of basic innovations. One basic innovation probably stimulates the emergence of additional ones, thus the clusters can be observed. During the growth phase, innovations are simply translated into improvements and do not act as catalysts for further innovative activities.

Nelson and Winter (1973, 1975 and 1977) extended the Schumpeterian description of economic growth as the aggregated behavior of individual firms in a competitive environment. While the work of Nelson and Winter does not deal directly with the long wave phenomenon, it can be easily extended to include the evolutionary technical change adopted by successful firms as a consequence of clustering of innovations. According to this evolutionary theory of decision making at the level of individual firms, continuous optimization is replaced by the maintenance of operation rules selected through competition until changes in the environment require a change in operation regime. The changes of the economic environment are in turn promoted by repeated selection of those firms whose decision rules are favorably adjusted to the prevailing conditions in the economy. Thus, the innovating entrepreneurs are the real promoters of economic change through competitive selection of the fittest. "The competitive environment within which firms operate is one of struggle and motion. It is a dynamic selection environment, not an equilibrium one. The essential forces of growth are innovation and selection, with augmentation of capital stocks more or less tied to the process" (Nelson, 1974). Consequently, Hartman and Wheeler (1979) extended this neo-Schumpeterian theory of evolutionary growth and technical change to include "the powerful impact of the clustering of innovations on patterns of firm growth and survival". They proposed "that waves of innovation and infrastructural development reflect at a macro level the concentration in space and time of the evolutionary technical change stipulated by Nelson and Winter" producing "as a result, major investment cycles and aggregate economic cycles" (Hartman and Wheeler, 1979).

3.2.2 Capital and Investments

According to Garvy (1943), Kondratieff used the theory of investment cycles of long duration to explain the long wave phenomenon. Essentially he borrowed from Marx the idea that business cycles are caused by the periodic reinvestment of fixed capital with an average life of about ten years, and introduced a graduation in durability in the production period and in the amount of investment corresponding to different kinds of capital goods. The replacement of basic capital goods that requires large investments over long periods, such as large production complexes, railways, canals and so on, takes place in spurts which due to the extended duration of projects cause the major cycles of economic life. Particularly large investments require large amounts of loanable capital. During a recession and depression the loanable capital becomes available at low rates due to the high propensity to save by those whose real income increases with the decreasing price levels. At this point reinvestment by entrepreneurial and financial groups begins and the inventions that accumulated during the recession and depression are put into practice. This explains both the clustering of innovations and the beginning of the recovery from depression. The expansion phase leads to prosperity that generates its own limitations. Hight interest rates and price increases both reduce the loanable capital and increase the rates at which it is available to entrepreneurs. Thus, Kondratieff explains the beginning of the recession period in terms close to those of the monetary over-investment theory.

Mandel (1964, 1980) extended the Marxist interpretation of the long wave as the successive acceleration and deceleration of capital accumulation and reinvestment. During recovery and prosperity both the volume of capital accumulation and the profit rate increase, while during the recession and depression the opposite is the case. Mandel contends that "it is possible, with the conceptual tools of Marxist economic analysis, to explain long-term upsurges in the average rate of profit at certain historical turning points, in spite of the cyclic downturn of that same rate of profit at the end of each industrial cycle, and in spite of the secular decline pointing to the historical limit of the capitalist mode of production" (Mandel, 1980). Thus, over the whole historical period of economic development the average profit rate portrays a declining secular trend, but its fluctuations around this trend are determined by the dynamics of the capital accumulation and the resulting long wave. Therefore, Mandel agrees in principle with Kondratieff's speculation that the postulated collapse of capitalism would bring about the disappearance of the long wave because the causes of the long wave are linked to the basic principles of the capitalist economy.

Mandel essentially describes the long wave in industrial capitalism development as succeeding stages of accelerated capital accumulation, overaccumulation, decelerated accumulation and underinvestment. These phases of the long wave are a consequence of cyclic movements of the prevailing profit rate - accelerated capital accumulation occurs as the profit rate rises during the recovery, overaccumulation is reached during prosperity because the average profit rate declines for the total accumulated capital stock, and recession and depression characterize the period of underinvestment because the profit rate reaches its lowest levels. At this point the introduction of basic innovations becomes possible because underinvestment "creates a historical reserve fund of capital, from which can be drawn the means for additional accumulation needed over and above 'normal' extended reproduction to allow a fundamental renewal of productive technology" (Mandel, 1980). The fall in the profit rate creates not only underinvestment, but it also makes capital available for renewal in form of promising investments. In time these investments lead to the acceptance of some of the technological and institutional innovations. This technological change offers new and increasing opportunities for profitable investments. Consequently, the profit rate rises and causes recovery and a new wave of capital accumulation.

The approach of Forrester et. al. (1983) in analyzing the causes and mechanisms of the long wave is based on system dynamics simulation. Essentially the whole economic development process is endogenous and depends only on the model specifications. Thus, Forrester does not require any direct empirical observation for the model, the dynamics of the model itself generate the long wave. The model contains two important mechanisms that create endogenous fluctuations. The multiplier-accelerator principle is employed to link investment to production and the capital accumulation adjustment mechanism which causes phases of capital shortage and abundance dúe to entrepreneurial perception and expectation lags. With appropriate parameter specifications, these basic relationships between production, demand and capital accumulation generate endogenously the fifty year cycle. It is interesting to note that Forrester's model does not need innovations to generate the long wave through renewal of production technology. Smooth demand growth is sufficient for the generation of the long wave. Nevertheless, Forrester recognizes the fact that clustering of innovations may occur before recovery, but he considers such clusters to be rather the effect of the favorable economic climate than the cause of the upswing. This neglect of the role of innovations in creating new technologies and growth sectors constitutes some of the criticisms of Forresters system dynamics model. It is probably not reasonable to assume that technological advancement is coincidental to the long wave and not one of its important causes. Ever since the industrial revolution each phase of unusually rapid growth and prosperity is associated with leading sectors that emerged out of an unusually large number of basic innovations.

3.2.3 Fluctuations in Employment

Freeman (1977) emphasizes the employment aspects of the long wave. Generally, the clustering of innovations and subsequent development of new industries and products increases the demand for labor and has favorable employment effects. Toward the end of prosperity and with the onset of recession and intensified competition, technological change would tend to generate labor displacing effects and unemployment would rise. Thus, during the recession, capital intensiveness of production increases as a result of increasing pressures toward labor and material savings. These measures become necessary in order to decrease costs as a response to intensifying competition. The initial growth of employment in new industries is also the cause of the eventual decline in employment once labor-saving technological change and the economies of scale from expanded production are achieved. While these employment fluctuations over the course of the long wave undoubtedly have important social implications, it is questionable whether they can be considered to be the primary causes of the long wave.

3.2.4 Availability of Raw Materials and Foodstuffs

Rostow (1975, 1980) emphasizes the causes of long waves in the growth of the world economy rather than individual countries. He distinguishes and relates three distinct causes of fluctuations in economic growth. The first unfolds as the result of the growth pattern of leading sectors with accelerated expansion stemming from the introduction and progressive diffusion of new technologies and eventual deceleration. The second cause is related to the changes in the profitability of "producing foodstuffs and raw materials, whether from the side of prices or technology, including their effects on investment in new territories and mines, on capital movements, interest rates, terms of trade, and domestic and international income distribution" (Rostow, 1975). The third cause results from large waves of international and domestic migration, or other changes in the rate of family formation, changes in housing and urban infrastructures and variations in the size of the working force. Rostow acknowledged Schumpeter's contribution in showing that during the Kondratieff downswing major technological innovations are a primary cause in generating price declines, rising real wages and thus narrowing real profit margins. He considers, however, that the early work of Kuznets on secondary movements in prices and production captures the powerful mechanism between price and production fluctuations in foodstuffs and raw materials. The sharp rise in relative prices of foodstuffs and raw materials at the end of the period of generally declining prices reflects the disequilibrium between supply and demand. Especially the growing population and rising real income due to general price declines are instrumental in increasing pressure on the food supply. The growth of leading industrial sectors creates equivalent difficulties in the supply of raw materials. With a certain adjustment process this relative price increase of foodstuffs and raw materials is translated in general price increases. As a remedy for relative scarcity and rising prices, new leading sectors develop that can provide abundant supply of foodstuffs and raw materials. Once they have taken the lead, the scarcity is effectively overcome and a new period of falling prices is initiated. This has a catalytic effect on the further diffusion of new sectors and eventually new limits are encountered and the cycle begins anew. Rostow's theory is essentially a price cycle, caused primarily by a shortage of primary inputs due to rapid economic and population expansion during an upswing. The relative scarcity then generates forces for introducing new sectors that can alleviate these pressures and lead to a new growth phase. The relative price level only reflects the succession of scarcity and abundance phases of world economic development.

4 LONG WAVES AND CHANGE OF TECHNOLOGY

The above brief review of the long wave theories has indicated that there are many potential indicators of the fluctuations that are reflected in the long wave. The emergence of basic innovations has a fundamental role in all theories, either as an endogenous necessity of the fluctuations or as exogenous factors that are related to the dynamics of economic development. Essentially all theories are complementary to the extent that they suggest the long wave consists of successive and alternating periods of growth and decline with similarity in terms of price movements, fluctuations in supply and demand, changes in capital accumulation and interest rates, unemployment rate, rate of technological progress, changes in innovative activity and so forth. The logistic substitution model has already illustrated that technological change and diffusion also follow regular patterns that have similar phases of successive alternation of growth and senescence following the introduction of a new technology. Each new technology grows to the limits imposed by the dynamics of the competitive behavior of other technologies sharing the same market, it then declines and with increasing obsolescence virtually disappears from the market. This process of technological rejuvenation is complementary to the technological and entrepreneurial innovative surges during the alternating phases of the long wave. The clustering of basic innovations is expected during the Kondratieff downswing. The empirical evidence for these clusters was speculated, among others, by Kondratieff and Schumpeter. Mensch provided the most exhaustive and convincing support of actual clusters around the years 1775, 1825, 1886 and 1935. Marchetti extended the analysis showing that inventions and subsequent innovation clusters are regular and recurring events. Therefore, it should also follow that market substitution of new technologies proceeded as a result of the emergence of these clusters after an appropriate gestation period and before actual commercialization and general acceptance of the subsequent technological development.

This hypothetical connection between technological substitution and the long wave is based on a conjecture derived from the phenomenological analysis of the logistic substitution patterns and the theoretical formulations of the major causes of the long wave in economic development. This hypothesis must be verified empirically before the exact nature of the two phenomena connected with the process of technological development can be related to each other. We will examine and document the evidence for the presence of long waves primarily in the economic development of the United Kingdom and the United States. This parallels the approach used in analyzing the dynamics of technological substitution in these two countries primarily because of their crucial importance as leading nations since the inception of the industrial revolution. The analysis will essentially consist of using a phenomenological approach to extract long fluctuations from historical records in an attempt to filter out the long waves and to compare the so-derived fluctuation patterns with the dynamics of technological substitution. The choice of this method is based on two assumptions. The first is that the success of the simple rules of logistic substitution in describing technological substitution led us to assume that equivalent, simple rules could be applied to distill long waves from price and "real" indicators. The second assumption is based on the belief that although every historical event or period is unique, some hidden invariants can be discovered with an appropriate method. "From the ubiquity of such events it follows that practically every economic fluctuation must be a historic individual and cannot be made amenable to explanation but by minute historical analysis of the innumerable factors actually at work in each case" (Schumpeter, 1935). Such a detailed analysis would be at least an overwhelming if not infeasible task. Consequently, we will apply simple rules to determine alternating periods in the economic past that can be identified and characterized in terms of invariants. In other words, we will analyze the unique sequence of historical developments with a long wave "filter" in order to discover invariant patterns of alternating periods. The postulated existence of invariants in the pattern of development could then be used as confirmation ("experiment") for a theoretical formulation.

4.1 Fluctuations and Secular Trends

Kondratieff (1926) and Schumpeter (1935) have already postulated the guiding principle for the search of invariants in the dynamics of long waves. They assumed that every sequence of annual economic (or other) quantities and indicators can in principle be decomposed into two components - a secular trend and the fluctuations around this trend. This may appear to be nothing more than a description of "curve fitting" with random (irregular) fluctuations. In this particular context this is not the case since we assume *a priori* that regular long fluctuations exist and attempt to confirm their occurrence with appropriate tools. In scientific research, a phenomenon can usually only be discovered if its existence is postulated. Here we will proceed with a neo-Schumpeterian definition of the secular trend and fluctuations.

By the term fluctuation we mean a periodic reoccurrence over the historical period under consideration of variation in values of indicators or quantities which do not display monotonic increase or decrease over time intervals longer than the duration of the periodic reoccurrence. Furthermore, we assume that these periodic reoccurrences do not occur independently in all time series, but rather that they display either instantaneous or lagged association with each other. Thus, we assume that a regular pattern exists between fluctuations of different time series. It is evident from this informal definition that some time series may portray only fluctuations. Others are in principle decomposable into fluctuations and a secular trend. By the term secular trend we mean that the historical period under consideration can be subdivided into shorter time-intervals for some time series in such a way that the mean values of indicators or quantities over these subintervals are either monotonically increasing or decreasing in time over the whole historic period or display recurrence only once. In addition we define that the periodic recurrence of fluctuations has a duration on the order of five decades. With these basic assumptions we have an heuristic "filter" that we can use to distill long waves from historical time series. Admittedly, these assumptions are perhaps too simplistic to be exactly encountered in statistical material, since obviously the empirical data contain innumerable other movements in addition to the fifty year fluctuations and secular trend. In this context, please note that we do not assume that other movements in the time series do not exist, but rather we choose to ignore them by making explicit assumptions about the fluctuations we would like to confirm. Otherwise, we would never be able to determine more than that empirical observations contain innumerable, superimposed random and recurring fluctuations. In other words, we have chosen a viewing filter for our observations that allows us to say more than that everything depends on everything else and that various movements cannot be untangled.

As an illustration of the proposed method for determining whether the long wave is a recurring phenomenon consider the Copernican model of planetary motion. When this model is used to process empirical data, the planetary motions are an extremely regular, recurring phenomenon that can be described with great precision by relatively simple rules. If the Ptolemaic approach is used, the recurring patterns are preserved, but the precision and simplicity in describing the planetary motions is lost. Although this example may appear to be somewhat presumptuous, we do not mean to imply that the dynamics of economic development display either the regularity of planetary motions nor that the relatively simplistic and preliminary character of most of the long wave theories should be compared with the Copernicus model and the almost universally accepted physical laws of planetary motions. Instead our intention was to illustrate the fact that certain, in principle complex phenomena, can be precisely described with simple rules if the appropriate hypothesis is posed to analyze the empirical observations.

In practical terms, our method consists of first eliminating the secular trend from non-stationary time series and then determining the residual fluctuations of the time series. The second stage consists of eliminating all other fluctuations shorter than the long wave. Usually, it is sufficient to form a moving average longer than the duration of the business cycle (i.e., longer than eleven years). This operation is not always necessary since the long wave movements are sometimes observable in the residual even without the elimination of shorter fluctuations.

Spectral analysis is a method for determining fluctuations of different periodicities in stationary time series. The advantage of this method is that it estimates simultaneously the contribution of all fluctuations with different periodicities to the total variance of the time series. Thus, it provides a method for determining the relative importance of different fluctuations without the need to eliminate shorter or longer fluctuations by smoothing from the series.

Unfortunately, the available time series usually cover a period only three to four times longer than the duration of the long wave. Typically, the spectral analysis could generate only three to four long waves compared with dozens of other cycles, so that the verification of the long wave would be difficult with this method. Its relative frequency is simply too low compared with other fluctuations. The disadvantage of the spectral analysis is that the series must be trend-free (i.e., stationary, see for example Anderson, 1971), an assumption that is rarely fulfilled. Thus, trend elimination is necessary, but the spectral analysis is very sensitive with respect to the method used.

In general, trend elimination from time series that are not stationary is usually more difficult than the decomposition of the stationary series into various fluctuations. Specifically, it is not always obvious which method of trend elimination should be used. We have used basically three different methods alternatively and in some cases we have applied more than one method for trend elimination in order to test the sensitivity of the obtained results with respect to such changes.

The first and simplest method is to form the geometric moving average for estimation of rapidly increasing or decreasing secular trends. The moving average must span sufficiently long time periods, at least fifty years, in order not to remove the long wave fluctuations. This procedure is otherwise very straight forward. The major disadvantage is that the estimated secular trends are based on fewer data points at the beginning and toward the end of the historical period under analysis. During the first four and last four decades of the period only moving averages shorter than fifty years can be formed.

The second method is applicable only to very rapidly increasing secular trends and involves the estimation of an exponential growth curve. This method is also relatively simple but is unfortunately not applicable to many cases that span historical periods of two hundred or more years. Besides the usual statistical tests for the goodness of fit, a very simple rough indication of the appropriateness of the exponential growth assumption for a particular time series is simply to plot the data on the logarithmic scale and check whether the secular trend forms a straight line or not. The major disadvantage of this method is basically that its applicability is limited to very few cases.

The third method involves the estimation of a logistic secular trend and this method can be applied to a wide range of examples. Its major disadvantage is that it is usually difficult to specify stable estimates for all the parameters of the logistic growth function from empirical data. The difficulty arises from the fact that the saturation level of a logistic growth process is not easy to determine if it has not already been reached during the observation period. In those few examples where this is the case the estimation procedure is simple. It reduces to the estimation of two parameters of a linear form, since every logistic growth process can be transformed into a linear form if the saturation level is *a prori* known. In the description of the logistic substitution model we have already mentioned that the theoretical saturation level is known - it is the ultimate market takeover. Thus, the fractional market shares of competing technologies were simply transformed into linear form and the estimation of the two unknown parameters did not pose any problems.

In the case of estimating the secular trend by a logistic function the procedure is substantially more difficult since we do not always know the saturation level and in some examples not even the lower asymptote, i.e., the theoretical value at the time when the growth process was initiated. In many examples this lowest value is different from zero and must be also estimated.

This is due to the fact that very few growth processes are initiated at infinitesimally low levels. Usually, the "start-up" of a growth occurs with some substantial level. As an example, consider the industrial production index for the world. At the dawn of human civilization the quantity of produced goods must have been negligible, but since recorded times the quantity was significant and larger than zero.

The problem in empirical estimation of a logistic growth process is that the saturation level, if it is not *a priori* known, is very sensitive to small changes in the data base. Bruckmann (1977) notes that sometimes the addition of a single new observation can change the estimate of the saturation level (i.e., the upper asymptote) significantly. He suggests that in such cases it is necessary to consider other factors, such as physical constraints, in order to obtain at least an interval for acceptable values. Bruckmann proposed two relatively simple methods for determining the initial estimates of the saturation level. Once this level is known with some precision an iterative procedure can be used to estimate the other two parameters and new saturation levels. We have used this procedure to obtain initial estimates for the parameters of the logistic function. We then use these estimates as initial values in a more complex estimation procedure. Variants of this procedure were developed by Oliver (1964, 1966 and 1969) and Nelder (1961). Essentially the method requires the solution of a set of four nonlinear simultaneous equations in the four parameters of the generalized form of the logistic function

$$\frac{g(t)-\gamma}{\gamma+\kappa-g(t)} = \exp(\alpha t+\beta),$$

where t is the independent variable usually representing some unit of time, and α, β, κ and γ are the four parameters of the function $g(t)$. κ is the upper and γ the lower asymptote of the generalized logistic function. We use an iterative procedure to solve this system of four equations with the initial values obtained from the other two simpler methods. A more detailed description of this and other estimation procedures is given in the Appendix.

Here we should also note that the logistic function is not the only S-shaped curve that is suited for the description of growth to limits. The four best known growth functions are the monomolecular, logistic (autocatalytic), the Gompertz curves (see Richards, 1959) and the integral of the normal distribution function. All four have been used to describe growth processes in a limited environment. The monomolecular function has no point of inflection, its growth-rate declines linearly with the level of achieved growth. Both the logistic function and the integral of the normal distribution are symmetrical round the point of inflection - the growth-rate increases before this point and decreases afterwards. The main disadvantage of the normal distribution is that the integral cannot be determined analytically, although the differential form is relatively easy to use. The Gompertz curve is asymmetrical around the inflection point. Thus, it can be only used for those growth processes where the assumption of such an asymmetry is justified. The logistic function is therefore better suited for description of S-shaped growth processes in spite of the difficulties associated with estimating its parameters.

4.2 Price Waves

The regularity of fluctuations in price data was the phenomenon that first stimulated Kondratieff and his predecessors to postulate the existence of long waves in economic development. These waves are most pronounced in the wholesale price indices for all commodities in the United States, but they can be observed in the price indices of other industrialized countries including the United Kingdom. Figure 4.1 shows the wholesale price index in the United Kingdom from 1560 to 1982 and Figure 4.2 in the United States from 1749 to 1982. In both countries prices appear to be stationary with long fluctuations almost over the whole historical period. Only after the 1940s can a pronounced inflationary trend be observed that had a magnitude greater than any other fluctuation before. In the United States prices reached pronounced peaks around the years 1780, 1815, 1865, 1920 and sharp increases during the last decade. In the United Kingdom, the fluctuations are subdued until the end of the eighteenth century. The first pronounced peak occurred around the year 1810, a weaker but prolonged peak around the year 1872, another pronounced peak in 1920 and a prolonged inflationary period during the last decade. Since all pronounced inflationary periods in both countries are associated with major wars, Hartman and Wheeler (1979) observed that the absence of strong inflation in the mid-nineteenth century United Kingdom could be partly due to the absence of such military conflict.

Clearly, price fluctuations in the United Kingdom and the United States display a broadly similar pattern, although the behavior in the United States prior to the mid-eighteenth century is not well documented and is uncertain. Prices in the United Kingdom portray a long decline from 1660 to about 1740 with two pronounced peaks in 1699 and 1710, and a long rise from about 1740 to about 1810, followed by another decline. The turning point between these two periods of rising and falling prices corresponds to the first pronounced peak. Especially large price rises occurred between 1785 and 1792 as the industrial revolution gained momentum. In the United States, a pronounced price peak occurred during the Revolution and recovery period between 1775 and 1785. Although the

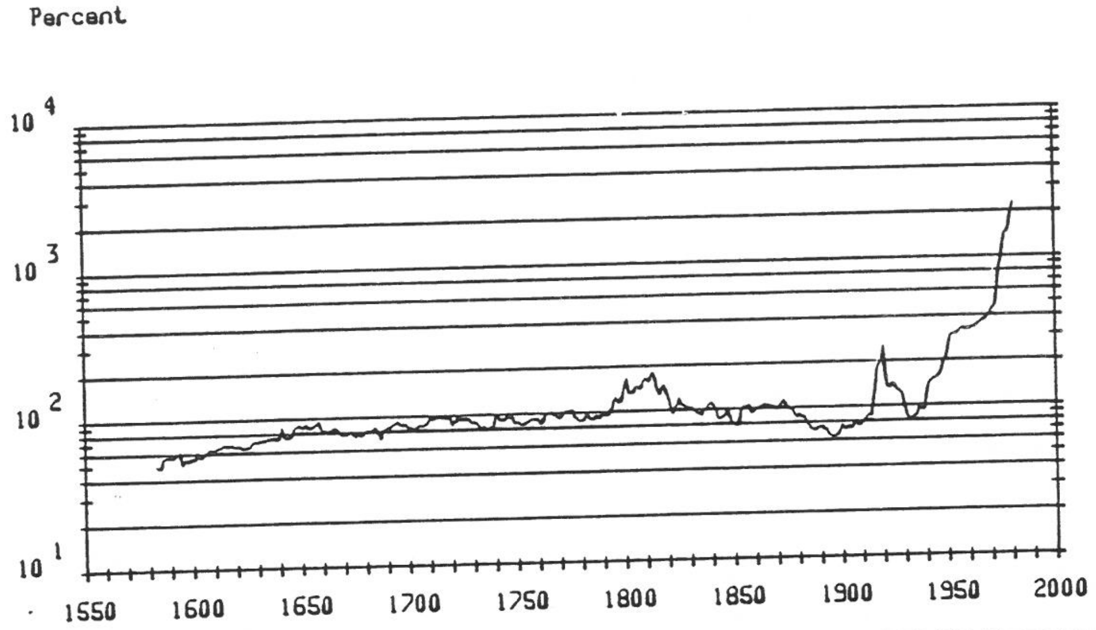


Figure 4.1 Wholesale Price Index, UK.

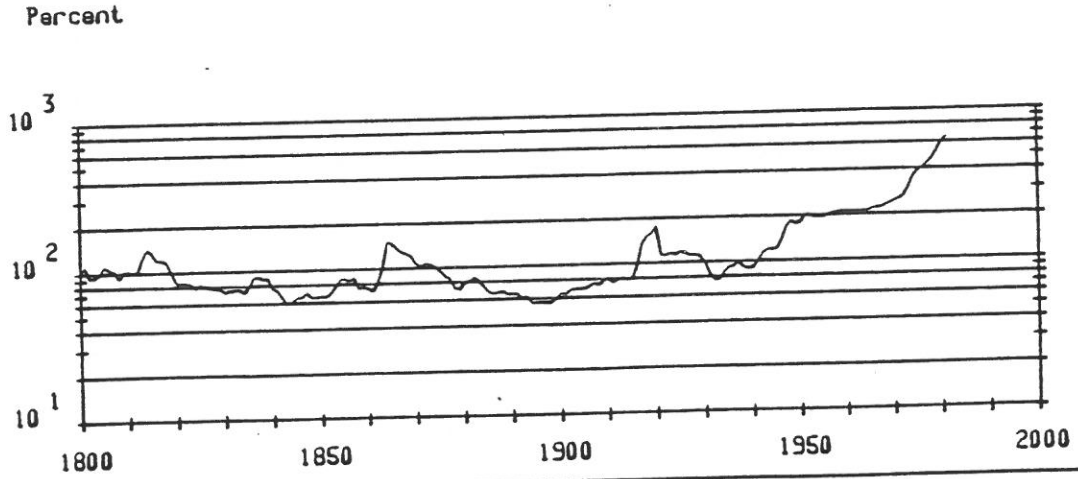


Figure 4.2 Wholesale Price Index, US.

two countries differed substantially in many respects such as the level of industrialization, institutional development, energy use, internal conflicts, etc., the parallel in the pattern of price fluctuations through the nineteenth and twentieth centuries is striking. In the face of quickened industrialization, the

Napoleonic Wars (UK) and the War of 1812 (US), prices rose until the 1820s. A period of declining trend continues through 1850 (UK) and 1843 (US), followed by a rising trend which is more pronounced in the United States, undoubtedly associated with the Civil War. The period from 1873 to 1896 is characterized by a declining trend in both countries and the succession of rising and falling periods has remained almost exactly parallel until the present. Since the Second World War, prices in the United Kingdom and the United States have risen almost uninterruptedly and to unprecedented levels.

The price fluctuations in both countries indicate a regular and parallel pattern as will be elaborated shortly. Price peaks of the 1780s, 1820s, 1870s, and the 1920s are spaced at intervals of four to five decades. These recurring long swings in prices are in our opinion not the primary causes of the long wave phenomenon but rather a good indicator of the succession of alternating phases of the long wave. A somewhat far-fetched analogy can be made if we consider the heart-beat as an indicator of the current state or level of activity of an individual. A high heart-beat rate and fluctuations would tend to indicate a rather high level of activity and general state of tension and alertness. A low heart-beat rate would indicate a subdued level of activity or regenerative state of deep sleep. In an analogous way we consider the long swings in price movements to indicate the phases of rapid growth and saturation with the increasing level of prices and phases of recession, and regenerative destruction with decreasing price levels.

In order to obtain a clearer picture of the timing of the long waves that are observable in the price indices of the two countries, we have decomposed the time series into fluctuations and a secular trend. Since the secular trend does not indicate a simple functional form we have used a fifty-one year moving average method for its elimination from the time series. We have smoothed the resulting residuals (i.e., the relative difference between the actual price level and its secular trend expressed as a percentage) with a fifteen year moving average. The resulting stationary series (smoothed and unsmoothed residuals) are shown on the lower plots in Figure 4.3 for the United Kingdom and in Figure 4.4 for the United States.

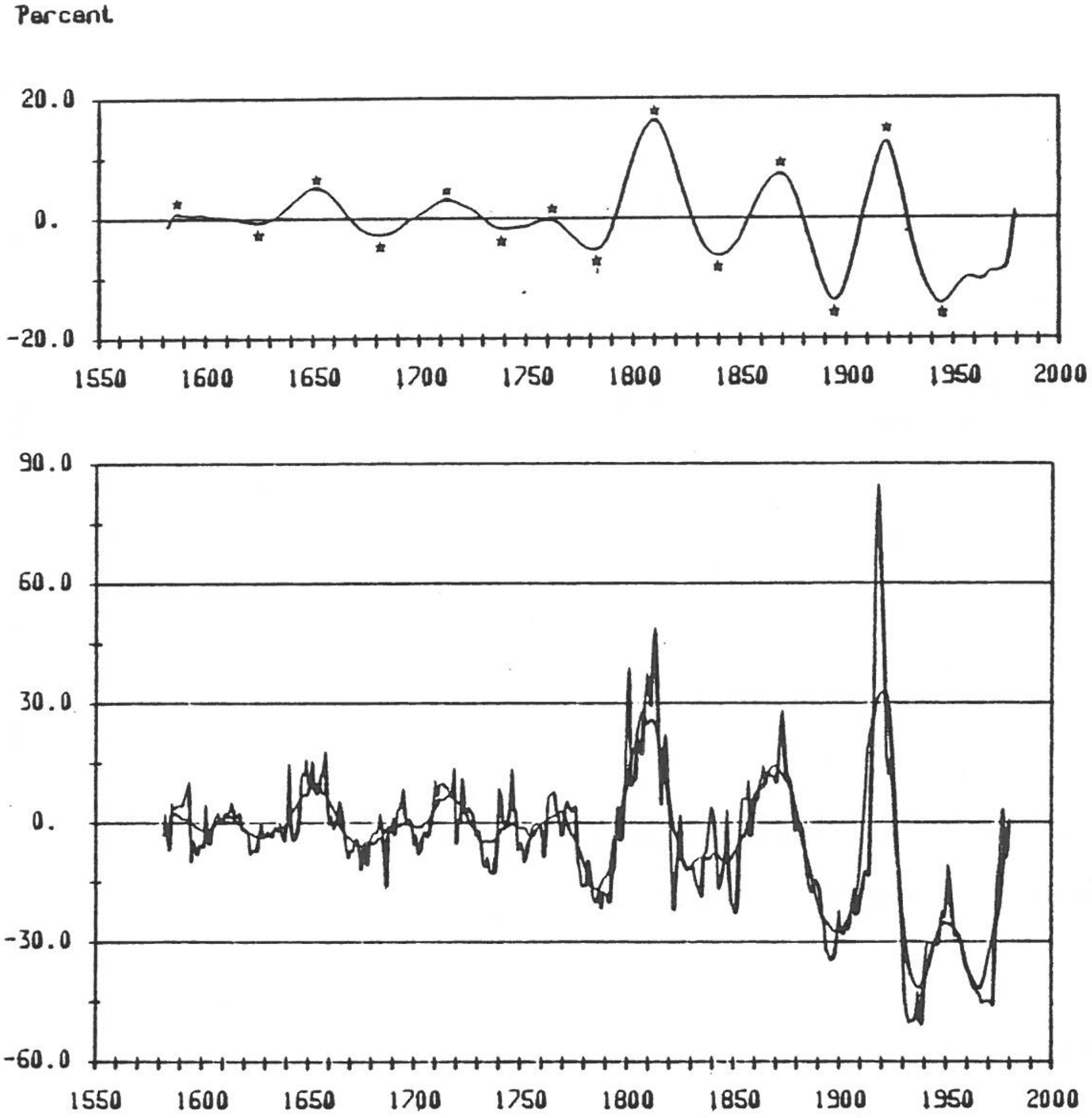


Figure 4.3 Long Wave in Wholesale Prices, UK.

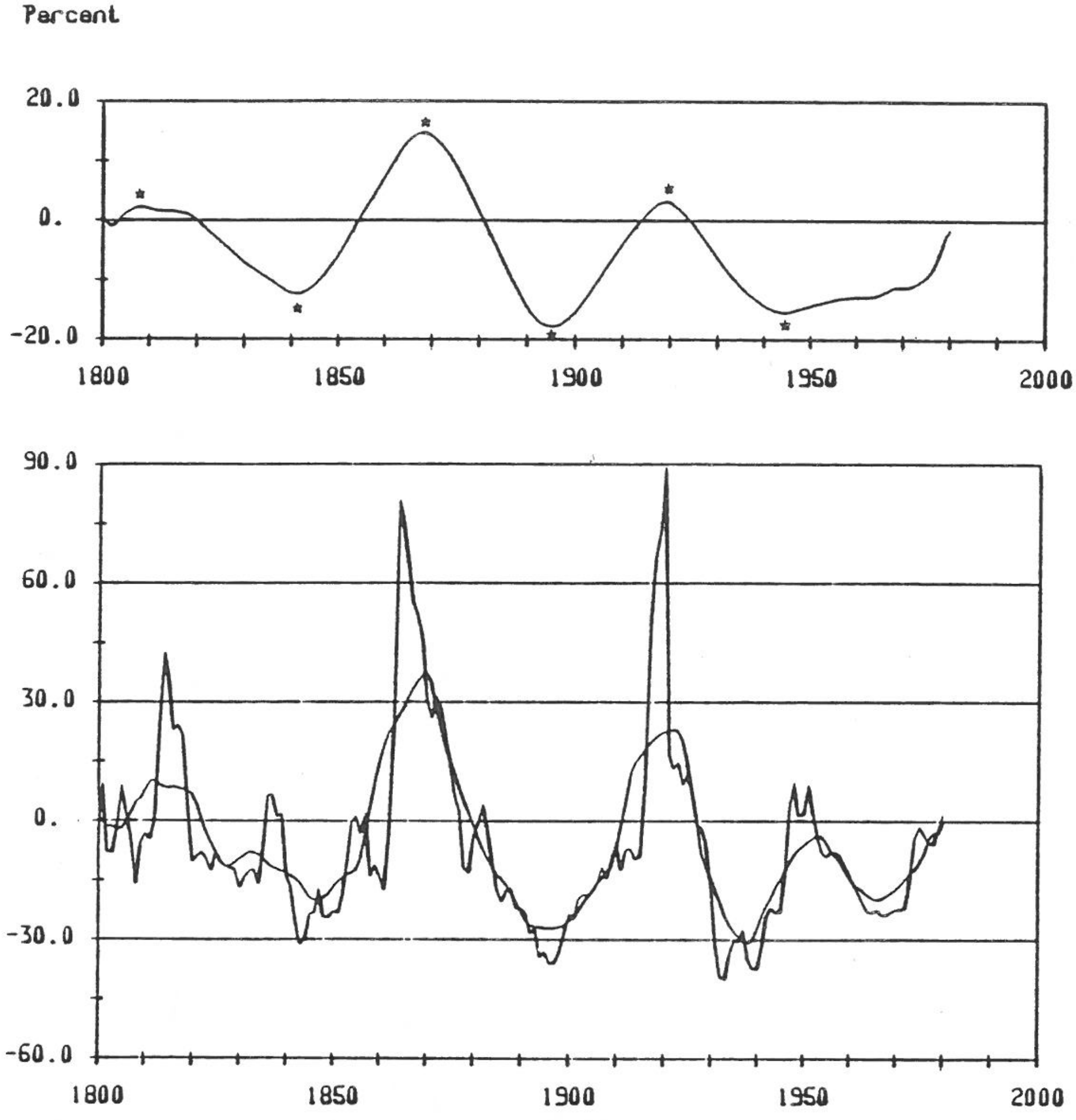


Figure 4.4 Long Wave in Wholesale Prices, US.

The upper plots in Figures 4.3 and 4.4 show *stylized indicators* of the long swings in prices. The curves have been derived by smoothing the residuals by a twenty-five (instead of a fifteen) year moving average. We have chosen such a long moving average in order to eliminate some of the more pronounced fluctuations from the residuals that overlap the five-decade long swings. The stars indicate approximately the turning points of the long waves.

This empirical determination indicates not only a parallel development of the price movements in the two countries, but also a high degree of synchronization. For the period before 1800 in the United Kingdom, we dated turning points in the years 1623, 1651, 1681, 1712, 1739, 1753 and 1773. The intervals between the four succeeding troughs are 58, 58 and 44 years and between the three peaks 61 and 41 years. As the industrial revolution gained momentum, first in the United Kingdom and later in the United States, the long swings became more regular and the magnitude of the fluctuations increased. The average amplitude of the fluctuations rose from less than ten percent in the pre-industrial United Kingdom to about twenty percent in both countries. Table 4.1 shows the dates of the turning points and the duration of the long swings in prices for the two countries. The average duration of the fluctuations is about fifty years and the occurrence of peaks and troughs varies by not more than one or two years. We consider Table 4.1 as a rough, empirical indicator of the timing of long waves in the two leading countries. In subsequent examples we will use this empirical indicator of the long wave turning points to determine the correspondence between the fluctuations that we will establish in other monetary and quantitative indicators of economic development.

Table 4.1 Chronology of the Long Wave, UK and US.

	Price Swings			
	United Kingdom		United States	
Phase	Period	Duration	Period	Duration
Downswing	1585-1623	48		
Upswing	1623-1651	28		
Downswing	1651-1681	30		
Upswing	1681-1712	31		
Downswing	1712-1739	27		
Upswing	1739-1753	24		
Downswing	1753-1773	20		
Upswing	1773-1810	37		
Downswing	1810-1840	30	1809-1841	32
Upswing	1840-1869	29	1841-1869	28
Downswing	1869-1895	26	1869-1895	26
Upswing	1895-1920	25	1895-1920	25
Downswing	1920-1945	25	1920-1945	25
Upswing	1945-		1945-	

4.3 Monetary Indicators

4.3.1 Purchasing Power of Gold

Gold is a rare monetary medium that can be measured and defined consistently over time. Since ancient times precious metals were used in exchange for goods and implicitly or explicitly formed the basis for currency. The gold price often mirrored major historical events and fluctuated simultaneously with major economic, political and military movements. The monetary role of gold and its purchasing power are crucial aspects and also indicators of historical events. In addition, a number of long wave explanations were connected with the relative scarcity or abundance of gold, especially during periods of unusually large discoveries.

Jastram (1977) argues that gold prices offer a standard of value for comparative purposes between different times, places and contexts. Over long time spans, gold provides one common way to compare the value of all forms of assets, fixed or liquid, and all kinds of income, earned or otherwise. Thus, the price and the purchasing power of gold provide a good historical yardstick for identifying different stages of economic development from growth to stagnation.

Figure 4.5 gives the gold price index, compiled by Jastram (1977), for the United Kingdom and Figure 4.6 for the United States. In both countries gold prices remained constant for long periods of time. The gold price index resembles a step function with few and relatively short interruptions during periods of currency debasement. In England the gold standard was established *de facto* in 1717 when the value of 21 shillings in money was tied to the value of gold in a guinea and not to the value of silver in 21 shilling denominations. Between 1560 and 1717 the gold price increased twice, during the 1610s and in 1662. This period of almost 160 years can be divided into three periods of constant gold prices (each of about 50 years) with the two price increases between them. Formally, Lord Liverpool's Act established gold as the sole currency standard 99 years later (in 1816) following the Napoleonic Wars. The first pronounced fluctuation of gold prices occurred during this period. From then on gold prices remained constant until the sharp peak around 1920 and the almost continuous price escalation from 1930 to the present.

In the United States gold prices remained constant from 1800 to the beginning of the Civil War with the exception of one small price increase during the 1830s. Following the Civil War the United States used inconvertible paper currency until 1879, when it resumed a *de facto* gold standard. In 1900, the Gold Standard Act formalized full convertibility of paper currency to gold. In 1933, the United States abolished the gold coin standard. Instead a new gold standard was established whereby the monetary units were to be defined in terms of gold, but

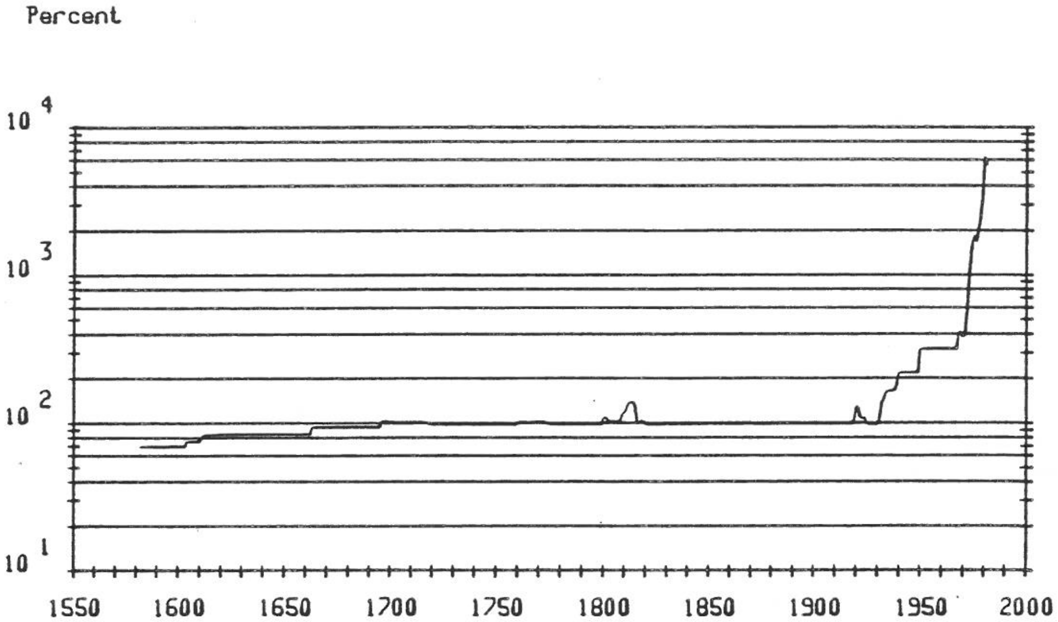


Figure 4.5 Gold Price Index, UK.

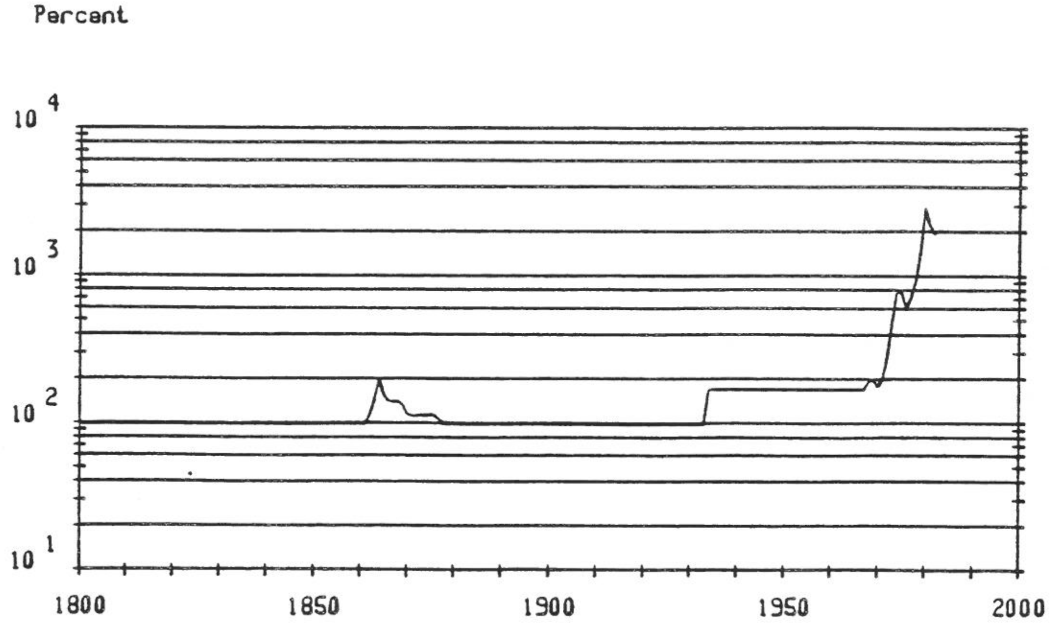


Figure 4.6 Gold Price Index, US.

the convertibility to gold was abolished. This change in the gold standard was also connected with a price increase. The price then remained constant until 1967, climbing to unprecedented levels afterwards.

The strong upward pressures on gold prices were encountered in 1961 when the London Gold Pool was established by eight of the most powerful central banks. Their aim was to stabilize the price of one fine ounce of gold to $35.0875. However, in 1965 the gold price rose slightly to $35.17 and the London Gold Pool begun to break up. Three years later the Pool was dissolved and a two-tier gold market emerged. Essentially, the United States established a monetary price of gold at $35 per fine ounce for official settlements between central banks. All other gold transactions were subject to free market exchange and due to the tremendous demand for gold on international markets the price surged to unprecedented levels until 1975. The trend reversed after 1975, but the price of gold is still more than ten times higher than before the *de facto* abolishment of the gold standard in 1968.

In spite of increases in gold prices the periods of constant prices were prolonged, lasting fifty or more years. In the United Kingdom, gold prices were essentially constant for over two hundred years between 1700 and 1920 with the exception of a pronounced peak during the Napoleonic Wars. During the historic period of relative constancy in gold prices given in Figures 4.5 and 4.6, Figures 4.1 and 4.2 indicate that the wholesale price indices in the two countries fluctuated crossing the path of gold prices repetitively. In other words, after long secular periods of inflationary and deflationary prices the value of gold remained essentially constant. This means that in the past it was possible to conserve the purchasing power of assets by converting them to gold and waiting until the next long swing in commodity prices. Thus, gold was a good hedge against price fluctuations provided one could wait long enough until the next price swing reestablished the original purchasing power of gold.

According to Jastram (1977), a more explicit indicator of the relative purchasing power of gold is the ratio of gold and commodity prices. Figure 4.7 and Figure 4.8 show this measure of the purchasing power of gold for the United Kingdom and United States, respectively, together with the wholesale price indices from Figures 4.1 and 4.2. Due to the relative constancy of gold prices over long time-periods, the purchasing power of gold and the price index portray a certain degree of symmetry. During the periods of constant gold prices the two indices are actually mirror images of each other, but due to the occasional increases of gold prices the secular movements of the two indices are not diametrically out of phase (i.e., the two indices are not mirror images of each other over the whole period). It is interesting to note that during inflationary periods (increasing prices) the purchasing power of gold weakens, while it increases during the deflationary periods. Since the periods of declining prices are associated with stagnation and depression, this means that the best time to convert assets into gold is at the pronounced price peaks at the end of the prolonged prosperity phase of the long wave. During the recovery and growth periods, on the other hand, the purchasing power of gold declines. Thus, gold is a good hedge against deflation and economic decline. In addition to this recurring conservation of gold value in exchange for other commodities after a price swing, a remarkable property of gold is that its purchasing power repeatedly returned to similar levels and its secular trend remained constant over the last four hundred years.

Figure 4.9 and Figure 4.10 show the long waves of the purchasing power of gold in the two countries, respectively, obtained by eliminating the almost

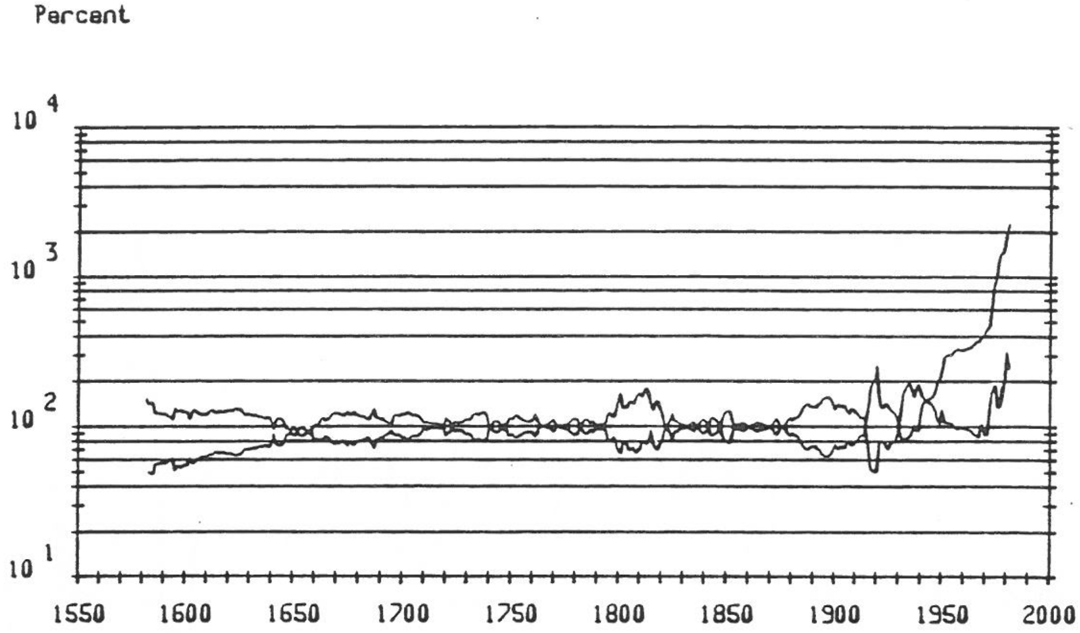


Figure 4.7 Purchasing Power of Gold and Wholesale Price Index, UK.

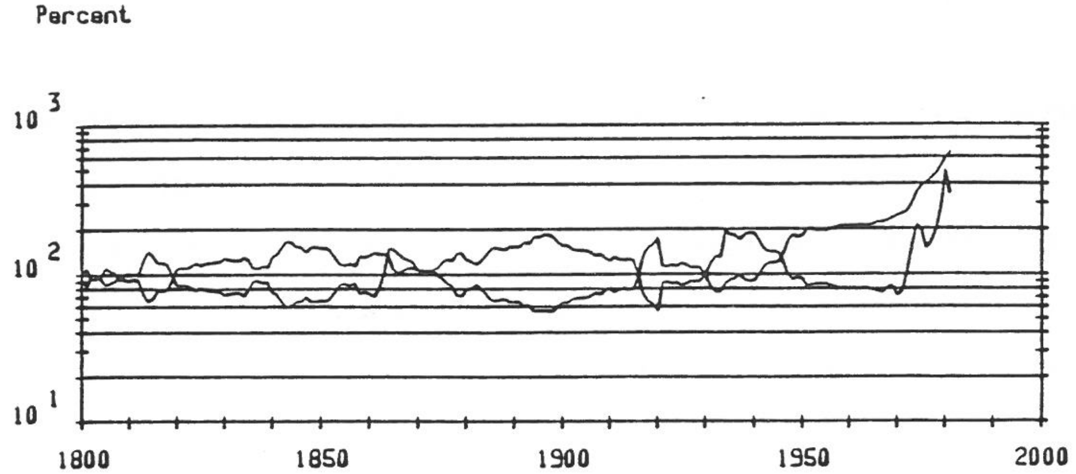


Figure 4.8 Purchasing Power of Gold and Wholesale Price Index, US.

constant secular trend from the time series and forming residuals. Both the annual fluctuations of the residuals and the smoothed fifteen-year moving averages are shown in the two figures. The purchasing power of gold displays regular long wave fluctuations - the upper turning points correspond to the

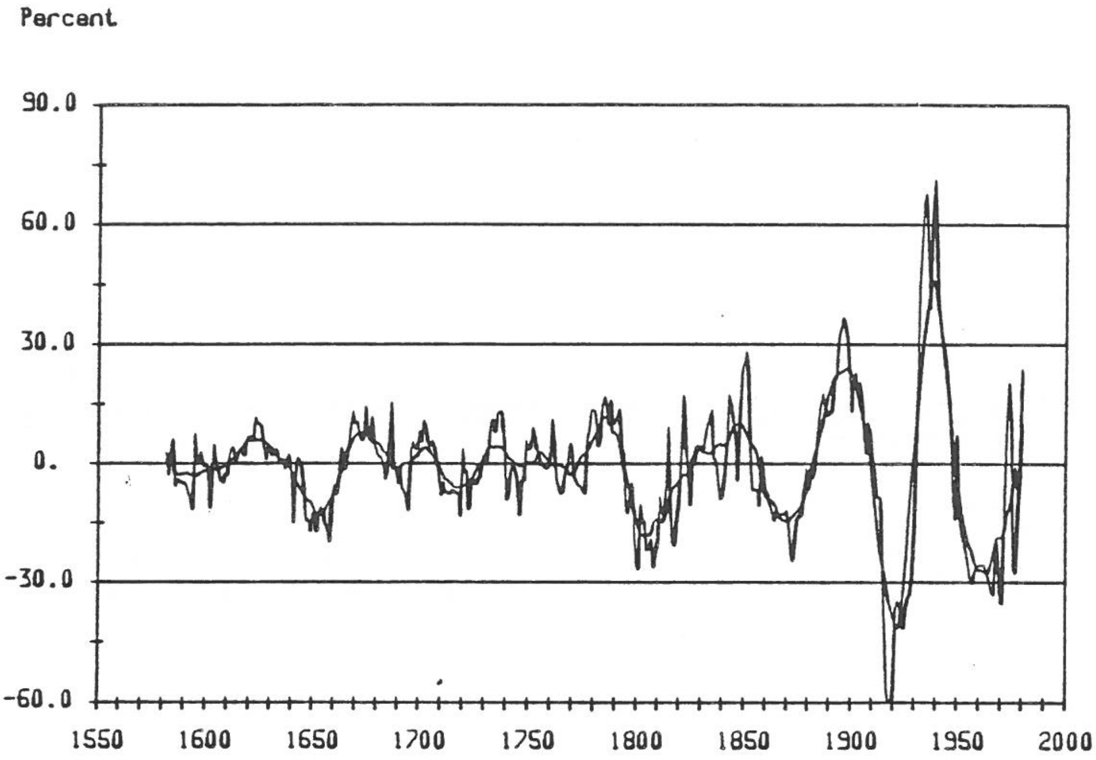


Figure 4.9 Long Wave in the Purchasing Power of Gold, UK.

depressions and the lower turning points to prosperity. Thus, the relative purchasing power of gold in exchange for other commodities is highest before the period of recovery. Consequently, the best strategy for long-term investment in gold is to keep liquid assets until deflation occurs and then convert to gold and sell when the first signs of inflation are felt.

4.3.2 Interest Rates

The prevailing rate of interest is another indicator of economic movements. Unfortunately, it is not easy to construct a good measure of the long-term fluctuations in interest rates. The problem is that in theory the true rate of

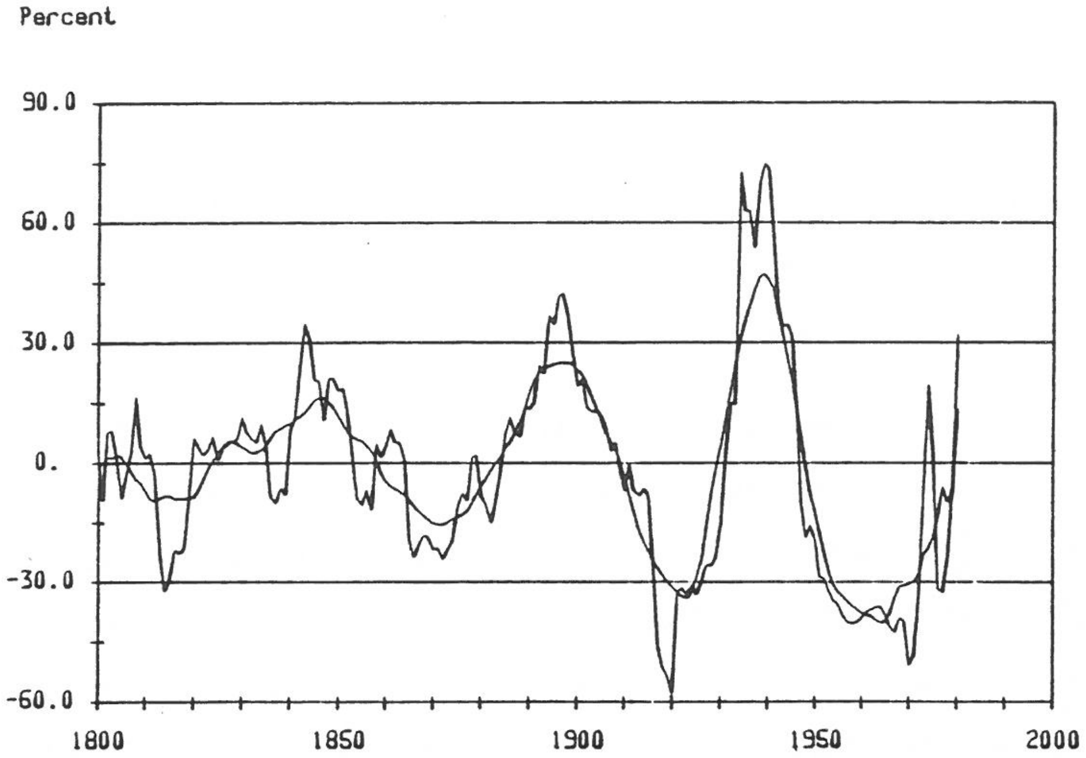


Figure 4.10 Long Wave in the Purchasing Power of Gold, US.

interest should be determined for loans of infinite duration without any risk of default and that it should not be affected by fluctuations in the value of money. In practice both of these conditions are not fulfilled. Loans have a finite duration to maturity and there is always some risk of default. We have seen that the value of money in terms of the gold standard and commodity prices did indeed fluctuate with the long waves of economic development. Perhaps the closest measure of the interest rate, with respect to its theoretical abstraction, is the times-series of the average annual flat yield of Consols in the United Kingdom since their inception in 1756 to the present. This series has the additional advantage that it is probably the longest unbroken record of interest rates available today (see, Mitchell and Deane, 1971). In spite of this favorable property, this series is not identical to the theoretical concept. Hicks (1957) noted that "it can hardly be maintained that at all points of this majestic sequence the Yield on Consols does satisfy these exacting conditions" of true interest rates. Although the Consols have traded now for over two hundred years their duration is not infinite. Mitchell and Deane (1971) observed that "for some years before 1888 the risk of termination by conversion existed, and the risk of default must have been felt at various times - for example in 1781, 1798, 1917, or 1940. Nevertheless, no better indicator of the long-term rate of interest exists".

Figure 4.11 shows the yield on Consols that can be divided into five periods. Until 1800 a secular rise can be observed followed by a prolonged secular decline until 1900. Another period of increase occurred up to the pronounced peak in 1920, followed by a short decline until 1948 and a final increase to unprecedented levels up to the present. The nominal yield on Consols has changed only three times during this long period. It was three percent per year from their first issue in 1756 to 1888. From 1889 to 1902 it was 2.75 percent per year, and from 1903 onwards it decreased to 2.5 percent per year.

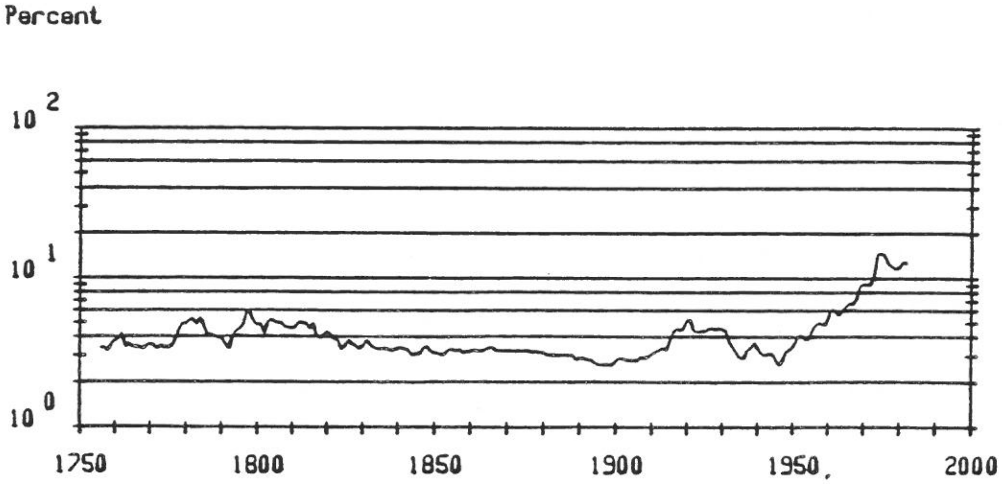


Figure 4.11 Yield on Consols, UK.

Figure 4.12 gives the smoothed and annual residuals obtained from the series after the secular trend was eliminated by a fifty-year moving average. The resulting residuals were smoothed by a fifteen-year moving average. The derived series indicates a high degree of synchronization with the long wave in prices. The upper turning points correspond to lower turning points in price movements and the lower turning points of interest rate fluctuations to the higher points for price movements. Thus, during prosperity phases prevailing interest rates reach lowest relative levels, while they reach peak levels during depression periods.

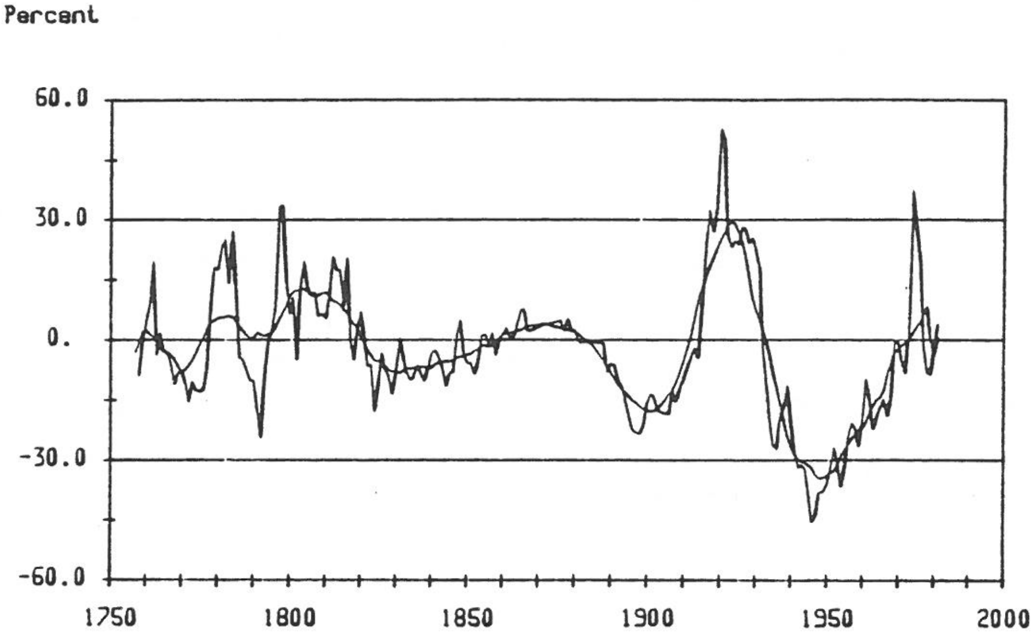


Figure 4.12 Long Wave in the Yield on Consols, UK.

4.3.3 Gross National Product

Gross national product series, despite all of their disadvantages, still offer the best aggregate indicator of national economic activity as a whole. In the

United Kingdom, attempts to estimate indicators of economic activity, such as the national income, date back to the end of the seventeenth century. The annual estimates of the gross national product are available starting in 1830. Figure 4.13 shows the deflated gross national product in constant prices.

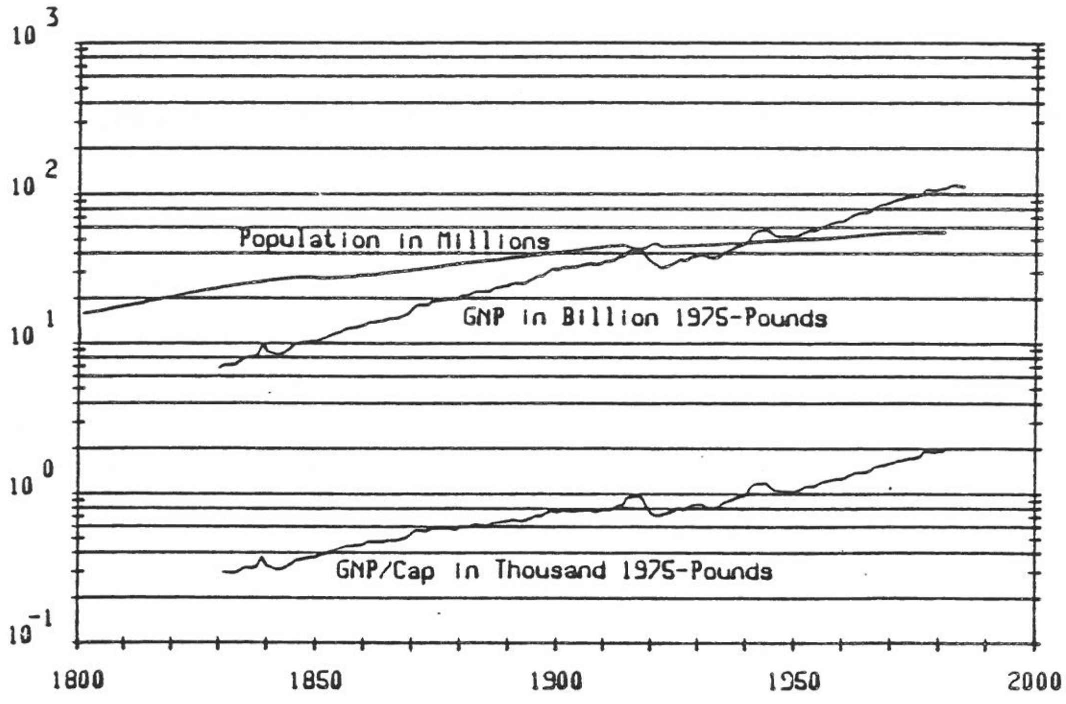


Figure 4.13 Gross National Product and Population, UK.

In the United States, the earliest available estimates of national income were compiled for the last decade of the eighteenth century. The annual estimates of the gross national product starting in 1890 were compiled by Kuznets (1961). Later, Berry (1978) revised the incomplete estimates available for some years of earlier periods and extended them to 1789. Figure 4.14 shows the resulting series of the deflated gross national product in constant prices for the United States. In both countries the secular trend portrays exponential increases with two pronounced peaks. In the United Kingdom the second peak also marks a change

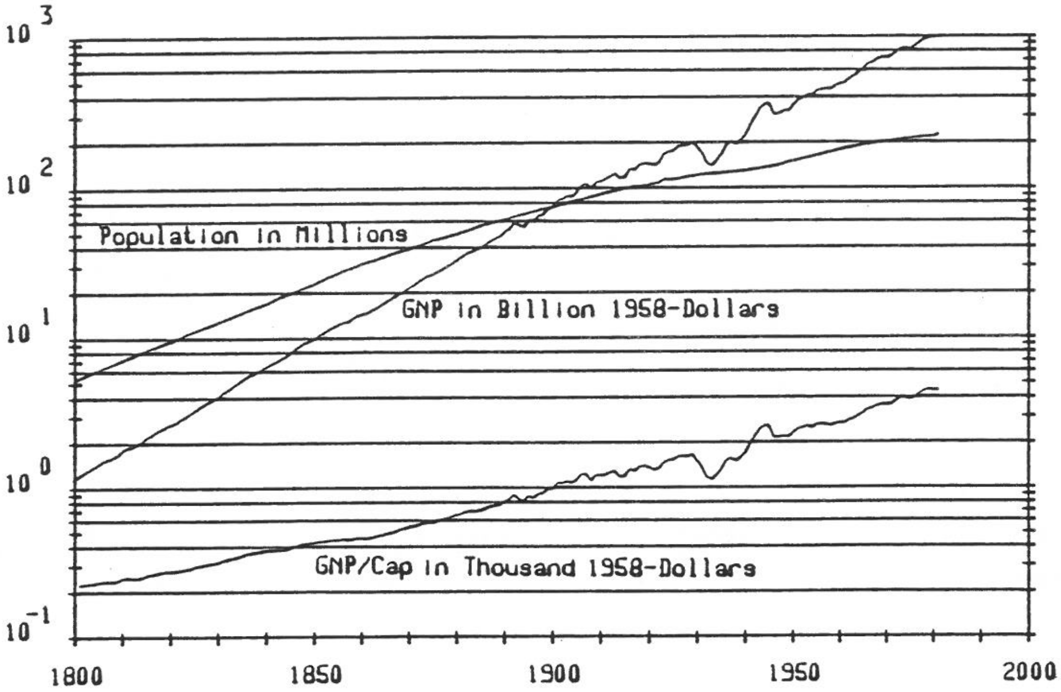


Figure 4.14 Gross National Product and Population, US.

in the secular trend to higher growth rates. In the United Kingdom the two peaks occurred in 1920 and 1945 and in the United States around 1930 and 1945. The smoothness of the estimates for the United States prior to 1890 indicates that the short-term variations are not captured in the estimates. Berry (1978) indicates that it is possible that his estimates come close to measuring the correct rate of long-term growth in spite of the lack of the typical short-term variations.

In order to eliminate a part of the secular growth from the two series, we have expressed gross national product in per capita terms before using the fifty-year geometric moving average to derive stationary residuals. The annual residuals and those smoothed with a fifteen-year moving average are shown in Figure 4.15 for the United Kingdom and in Figure 4.16 for the United States, respectively.

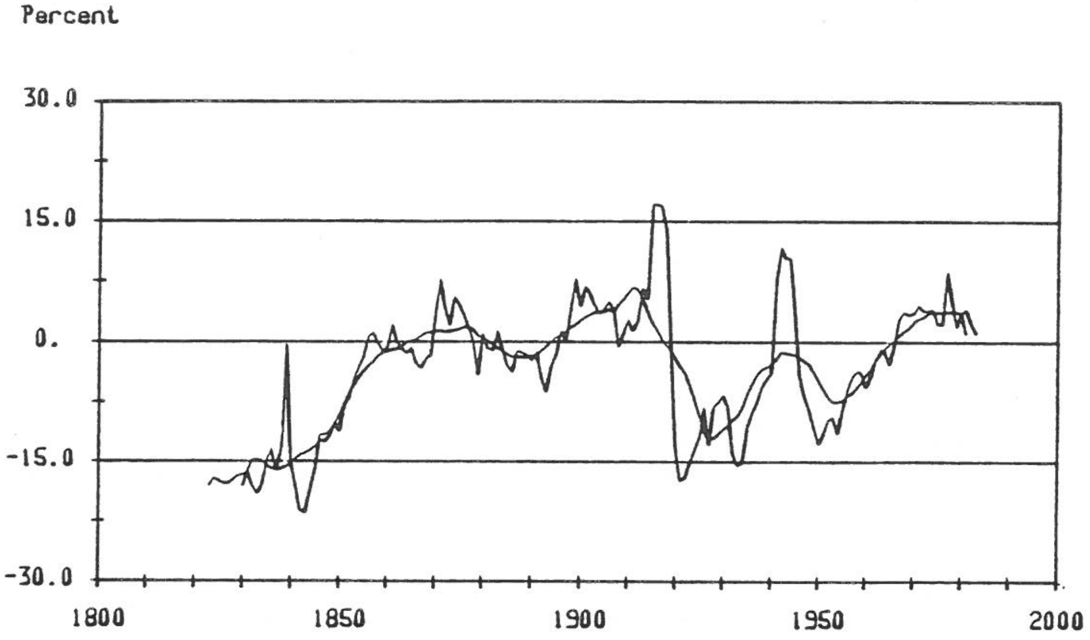


Figure 4.15 Long Wave in Gross National Product, UK.

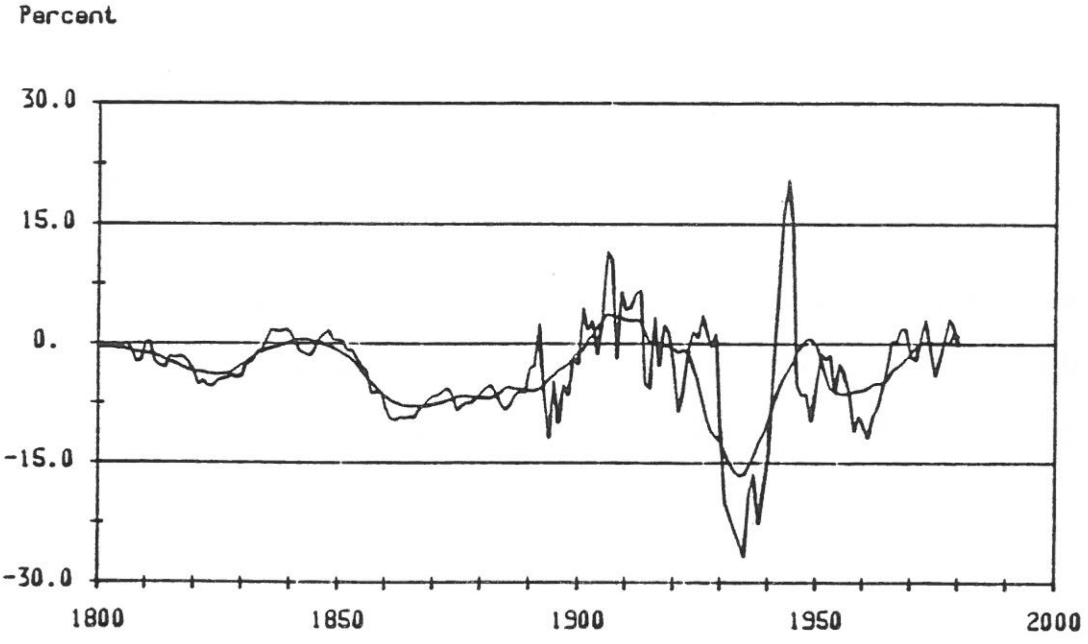


Figure 4.16 Long Wave in Gross National Product, US.

The turning points of the resulting long-term fluctuations parallel those observed for price fluctuations. The amplitudes of the movements are less regular than those of the price indices. During the last lower turning point, the price fluctuations indicated a weaker secondary cycle between the 1940s and the 1960s (see Figures 4.3 and 4.4). In the case of gross national product this secondary movement is even more pronounced, although the magnitude of the fluctuation is not as large as the amplitude of the last long wave. Nevertheless, the gross national product series do not portray the same regularity of long wave movements as prices or the purchasing power of gold. This could in part be accounted for by the difficulties of estimating the actual levels of overall economic activity over such long periods from the records that are available.

4.4 Energy Consumption and Production

Energy use is one of the rare quantitative indicators that can, at least in principle, be compared over long periods of time in spite of many technological changes and substitutions of old by new sources of energy. This is possible because the use of different energy sources can be expressed in common energy units. A kilowatt-year of primary energy consumed two centuries ago expresses the same magnitude as the same unit of energy consumed today, in spite of the great difference in the properties and qualities of the energy forms that were actually used at these widely separated points in time. The major difficulty associated with such comparisons is that most of the energy used during the early periods of the industrial revolution constituted non-commercial sources. We have already discussed the problems involved in estimating the levels of non-commercial use in the past (see Chapter 2). In the context of long waves we are only interested in relative changes in the levels of energy use and not in the growth and relative shares of various energy sources. These fluctuations around the secular trends, however, may be to an extent marked by the fact that especially fuelwood, the most important of all non-commercial energy sources, was estimated primarily on the basis of per-capita use. Thus, since the fuelwood time series do not represent actual use, but rather serve as an indicator of the relative importance of its use, some of the fluctuations may not be contained in the data. In spite of this drawback we will first consider primary energy consumption and than consumption of individual energy sources.

In Figures 2.3 and 2.10 we have already shown primary energy consumption in the United Kingdom and the United States, respectively. We have noted there that at least three distinct phases in the growth of energy consumption can be observed in both countries. In the United Kingdom, the first phase ends in 1800 followed by a more rapid increase until 1920, and then by a relative decline and rapid growth during the last three decades with a renewed decline in the last few years. In the United States, a phase of more rapid growth starts in 1900 and continues until 1930. After a short interruption the rapid growth resumes a few years later and continues until the last few years.

Figure 4.17 and Figure 4.18 show the annual fluctuations of energy consumption (from Figures 2.3 and 2.10) resulting after the elimination of the secular trend by the geometric fifty-year moving average trend together with the same fluctuations smoothed with a fifteen-year moving average for the United Kingdom and the United States, respectively. In both countries, the fluctuations show the same regular and parallel movements as the long waves in prices (see Figures 4.3 and 4.4). In the case of the United Kingdom, a major departure can be

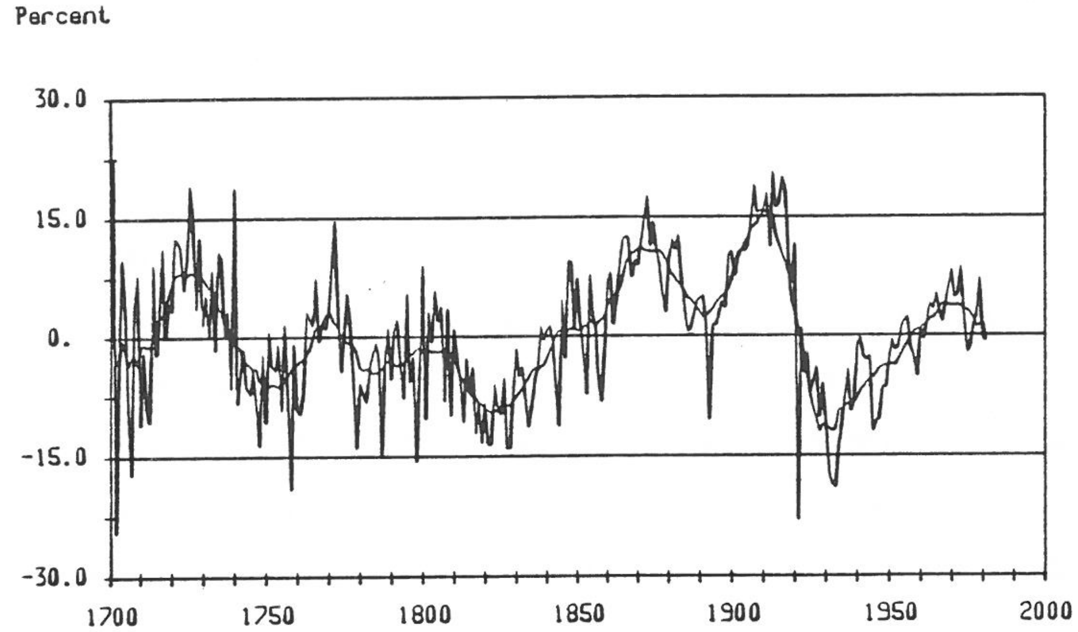


Figure 4.17 Long Wave in Primary Energy Consumption, UK.

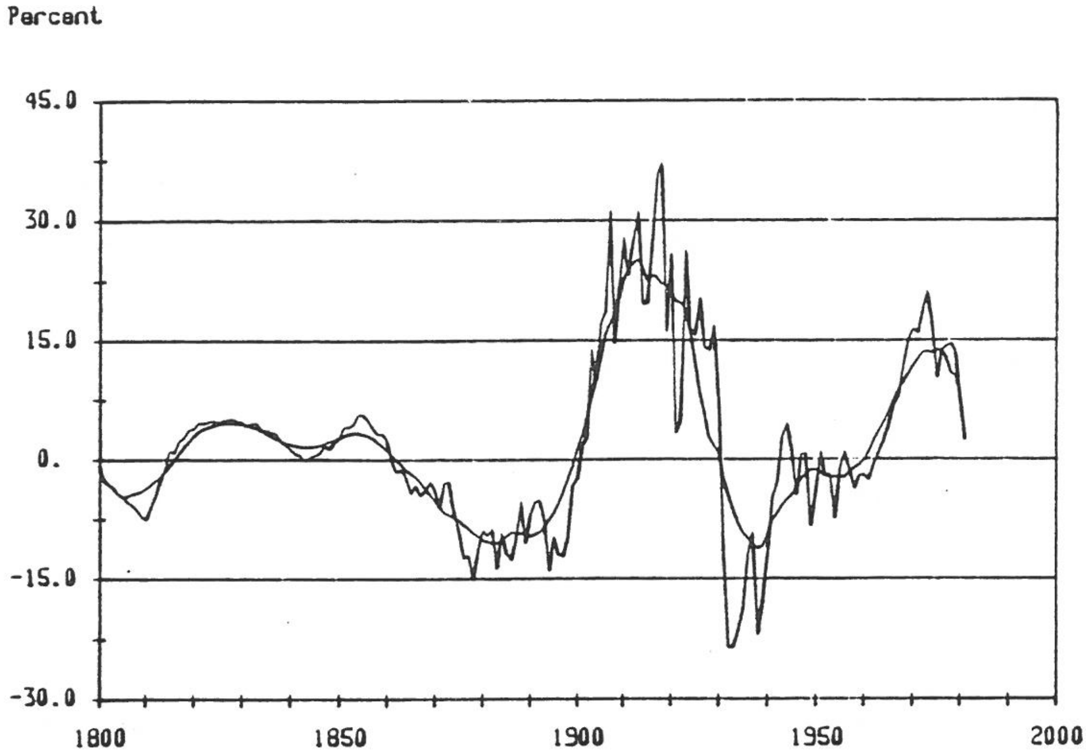


Figure 4.18 Long Wave in Primary Energy Consumption, US.

observed at the third lower turning point. This occurred in 1820, seventeen years earlier than in the price movements, however this difference was already reabsorbed during the next wave. The next upswing in energy consumption also shows a pronounced secondary cycle during the 1840s that is not present in the long wave in prices. In the United States, the second upper turning point in energy consumption is not as pronounced as in price movements and it also occurred somewhat earlier - in 1855 instead of 1870. This could in part be explained by the fact that about ninety percent of total energy consumption was supplied by fuelwood during this period (see Figure 2.11). As was mentioned earlier, the fuelwood time series represent estimates that were based primarily on the population growth so that they do not portray many fluctuations observed in other energy sources.

4.4.1 Analysis of Secular Trends

The secular trend of primary energy use in the United States can be captured by a number of functional forms. Stewart (1981) used the logistic growth curve to eliminate the secular trend basing his estimate on five-year averages of primary energy consumption. The resulting fluctuations around this trend showed pronounced long waves. The drawback of this approach is that he used shorter time series starting in 1860, so that only the last and the current wave were displayed. Our data base goes back to 1800 and extends over one more wave.

In order to investigate the sensitivity of the long wave in primary energy consumption, we will use our extended data base and will employ three different estimation methods of the secular trend. We have already showed the example obtained by using the geometric fifty-year moving average. Here, we will compare this method with the logistic and exponential growth curves. Both of these methods can be applied to the data set since the (unknown) saturation point of primary energy consumption has not yet been reached in the United States. There are many reasons to assume that a saturation, or at least lower growth rates, of primary energy consumption will occur in the future. Perhaps the most trivial of them, but also a very convincing reason, is that every growth process eventually reaches some limits. On the other side, if one only considers the evolution of primary energy use in the United States on the basis of growth rates displayed during the last 180 years, then the evidence for assuming constant growth rates is very strong.

Figure 4.19 shows the historical primary energy consumption in the United States with two alternative secular trends. The first is a four-parameter logistic fit with a saturation level of about eight TWyr/yr to be reached after the year 2050. The second estimate is an exponential fit that leads to astronomical consumption levels in the far future. The derived annual growth rates of primary energy consumption over the historic period are invariant to the estimation method used at about three percent per year.

Figure 4.20 shows the residuals, smoothed with a fifteen-year moving average, resulting from the three alternative estimation methods of the secular trend (the logistic and exponential estimates and the fifty-year geometric moving average from Figure 4.18). The turning points of the fluctuations are almost invariant to the estimation method. The amplitudes of the fluctuations change depending on the estimation method. Especially the amplitude of the last upper turning point in 1975 is very sensitive. It is lowest in the case of the exponential fit since the low rates of energy growth during the last ten years are below the

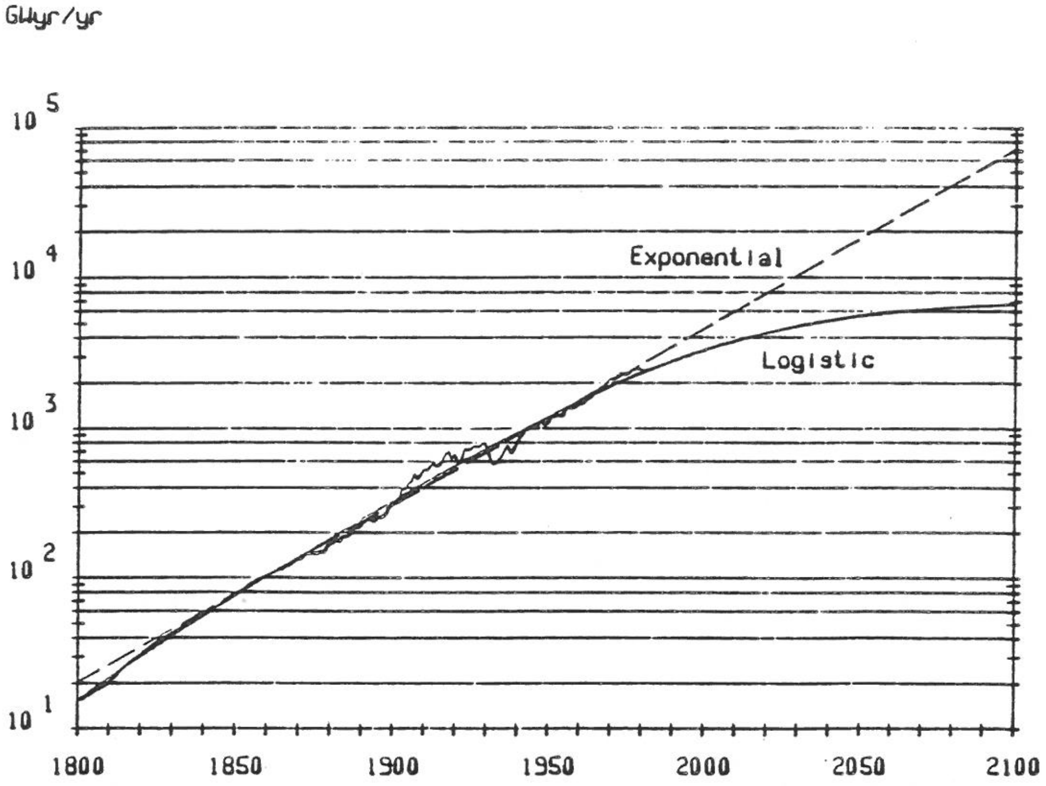


Figure 4.19 Primary Energy Consumption, US (with Two Secular Trends).

trend of the exponential growth curve. It is also interesting to note that the lower turning point of the first wave in Figure 4.20 is dated in 1897 by the moving average method and in 1883 in the case of the exponential and logistic methods. In spite of such small changes in the dating of this turning point and a larger variance in the amplitude of the last wave, the parallel fluctuations of all three long wave curves indicate that the broad features of the fluctuations in primary energy consumption are not a function of the method used to eliminate the secular trend from the data. Apparently, all three methods are suited for trend elimination in this particular context, and since the moving average is the easiest to compute, this sensitivity analysis offers an *a posteriori* justification for using the simplest method of trend elimination in most examples.

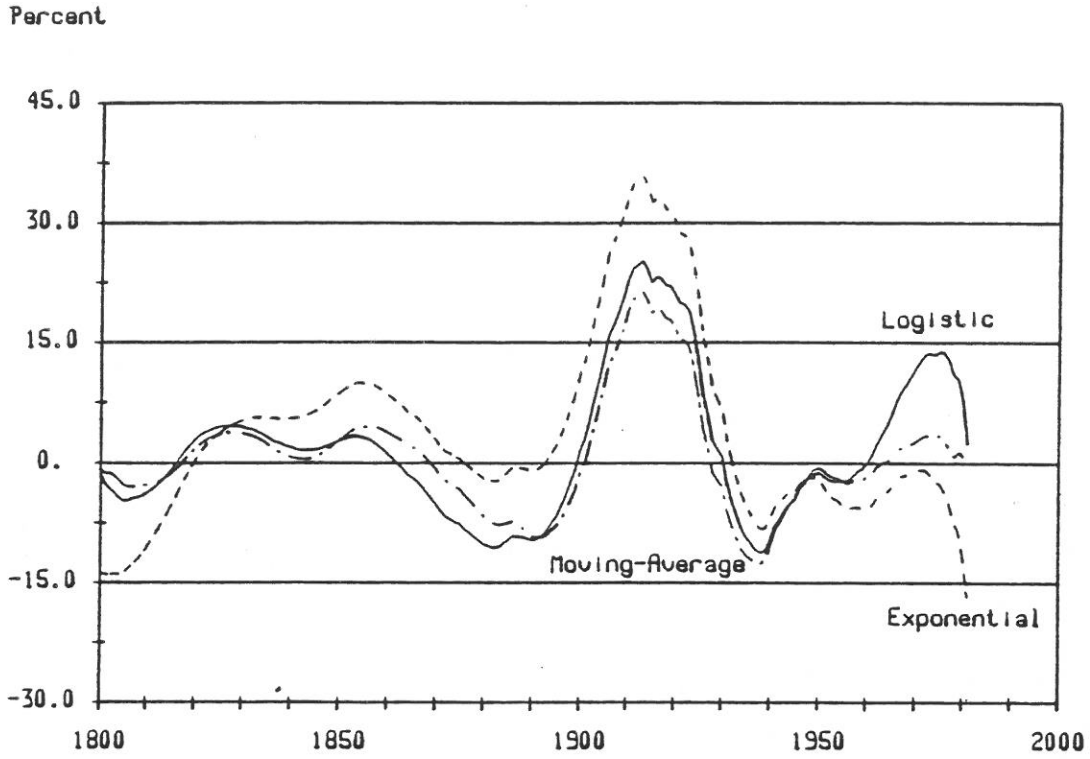


Figure 4.20 Long Wave in Primary Energy Consumption, US (Three Estimates).

4.4.2 Fossil Energy Use

We have seen that primary energy consumption as such portrays long wave movements. Here we will investigate the long-term fluctuations in the consumption of individual energy sources. It is conceivable that the rapid growth of new energy sources, as they undergo a vigorous market penetration and substitute the older sources, would tend to introduce additional fluctuations. In addition, the consumption levels of fossil energy sources are known with greater certainty than the estimates of older, non-commercial energy sources. This is especially critical in the United States where fuelwood constituted the major source of energy during the last century. In the United Kingdom, on the other hand, coal constituted the major source of energy since the seventeenth century all the way up to the 1950s. Here, it is interesting to consider the behavior of coal use separately from crude oil in search of different fluctuation patterns.

Figure 4.21 shows the trend-eliminated fluctuations of coal consumption in the United Kingdom and Figure 4.22 the fluctuations in coal production. Both figures were obtained by first eliminating the secular trend by a fifty-year geometric moving average and then smoothing the residuals with a fifteen-year moving average. Coal consumption and production curves portray the same fluctuations as primary energy consumption from Figure 4.17 up to the end of the last century. This is of course not surprising since most of the energy consumed during this time was provided by domestic coal. During the last eighty years, the fluctuations are different because rapid expansion of oil use introduced new variations in the overall energy consumption patterns and because the United Kingdom started importing coal from abroad. Nevertheless, it is interesting to note that the turning points of the fluctuations are the same in all cases up to the 1950s. At this point, as crude oil became the major source of energy, the upper turning point of primary energy consumption is delayed about ten years (1970) when compared with coal consumption and production (1960).

Figure 4.23 shows the fluctuations in crude oil consumption that were obtained by the same method, i.e., the secular trend was eliminated by a fifty-year geometric moving average and residuals were smoothed by a fifteen-year moving average. This figure illustrates that the delay in the last, upper turning point is no doubt due to the fluctuation in oil consumption. In fact, the oil consumption long wave is out of phase with the long wave of coal consumption and production. As oil became the major source of energy during the 1950s it also changed the pace of the long wave in energy consumption.

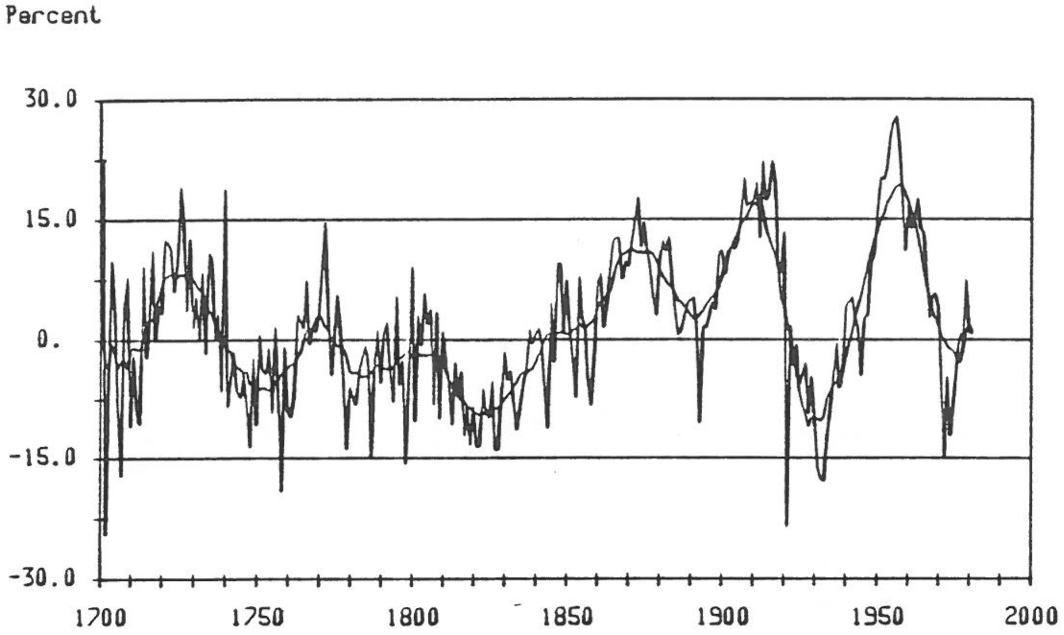


Figure 4.21 Long Wave in Coal Consumption, UK.

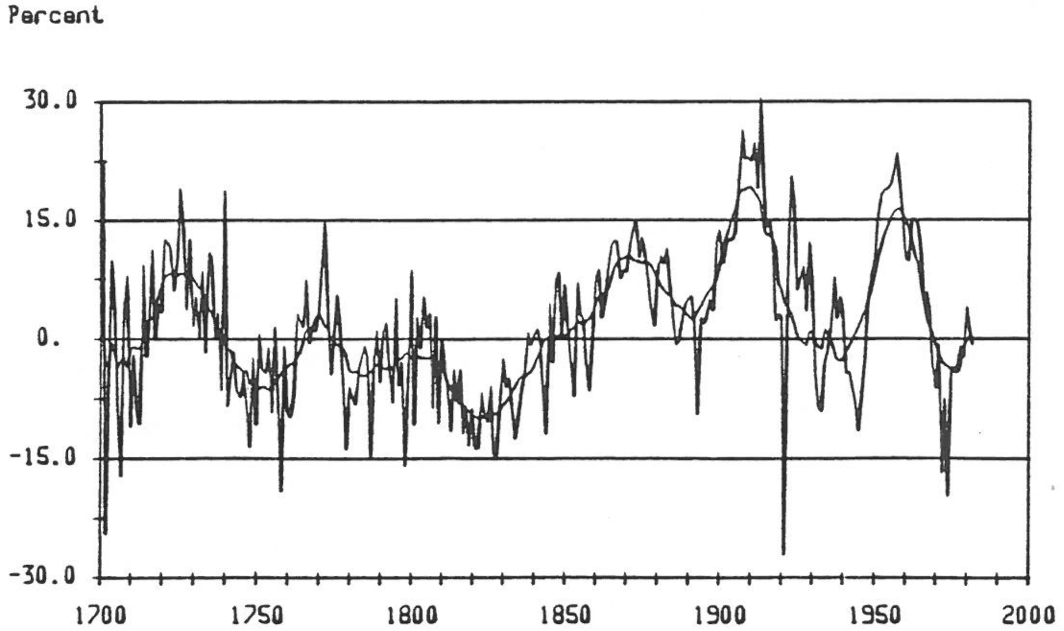


Figure 4.22 Long Wave in Coal Production, UK.

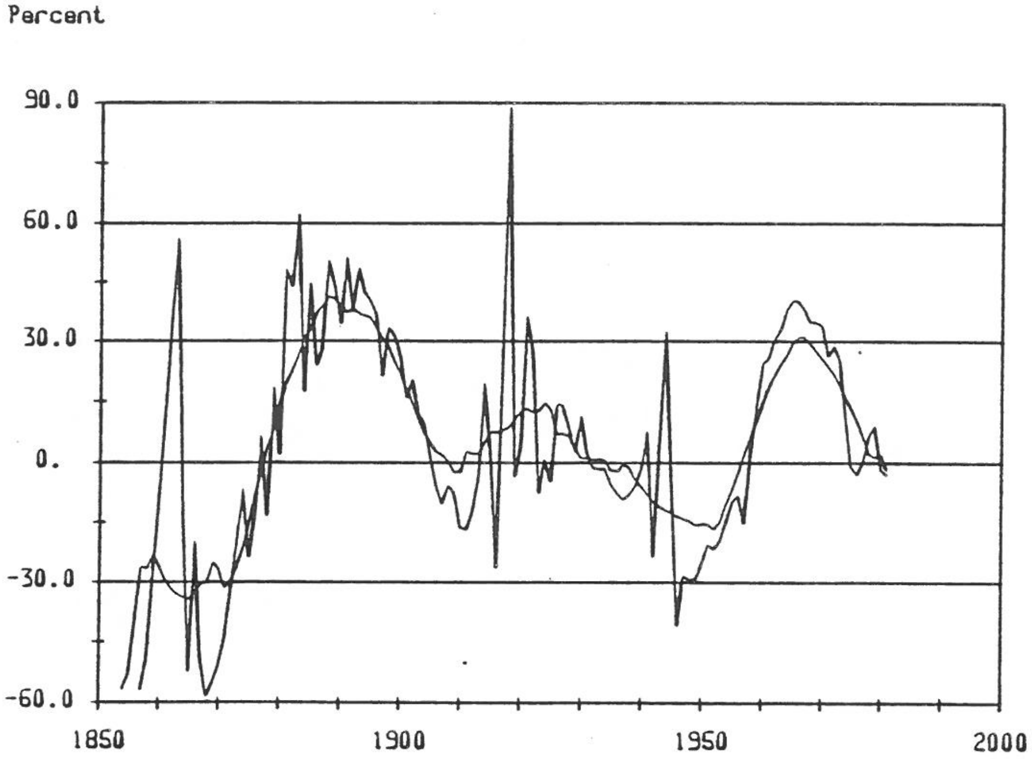


Figure 4.23 Long Wave in Oil Consumption, UK.

The same behavior can be observed in the United States, although it is not as pronounced as in the United Kingdom. Figure 4.24 shows the fluctuations in fossil energy use (i.e., fuelwood was eliminated from the data set). The fluctuations indicate the long wave more clearly than the total primary energy consumption from Figure 4.18. Fuelwood consumption (see Figure 2.10) is very smooth, probably because population growth (see Figure 4.14) was one of the most important secular trends used to estimate the data. Thus, during the last century when fuelwood was the most important source of energy, it obscured some of the fluctuations present in fossil energy sources. Figure 4.25 and Figure 4.26 show the fluctuations in coal and oil consumption, respectively. The comparison of these two figures shows clearly that the fluctuations are out of phase. However, the amplitude of oil variations around the secular trend is not

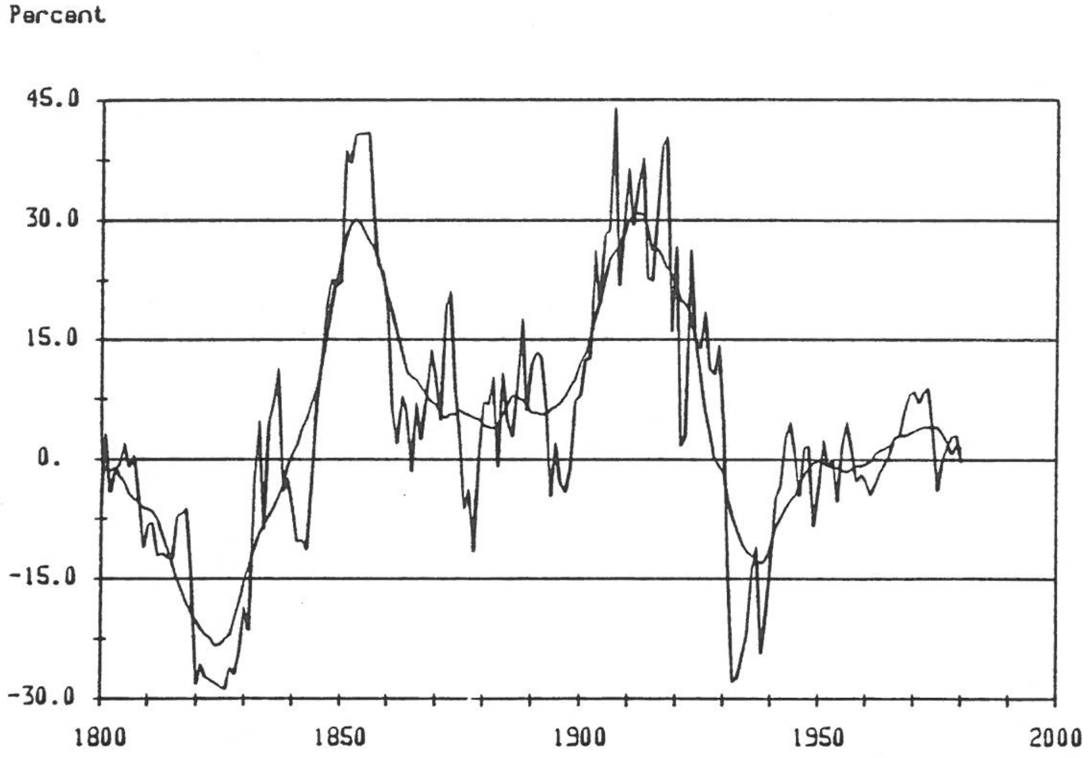


Figure 4.24 Long Wave in Fossil Energy Consumption, US.

very large indicating that the initial increases in oil consumption were limited more by other factors that may be independent of the long wave fluctuations observed in older energy technologies such as coal.

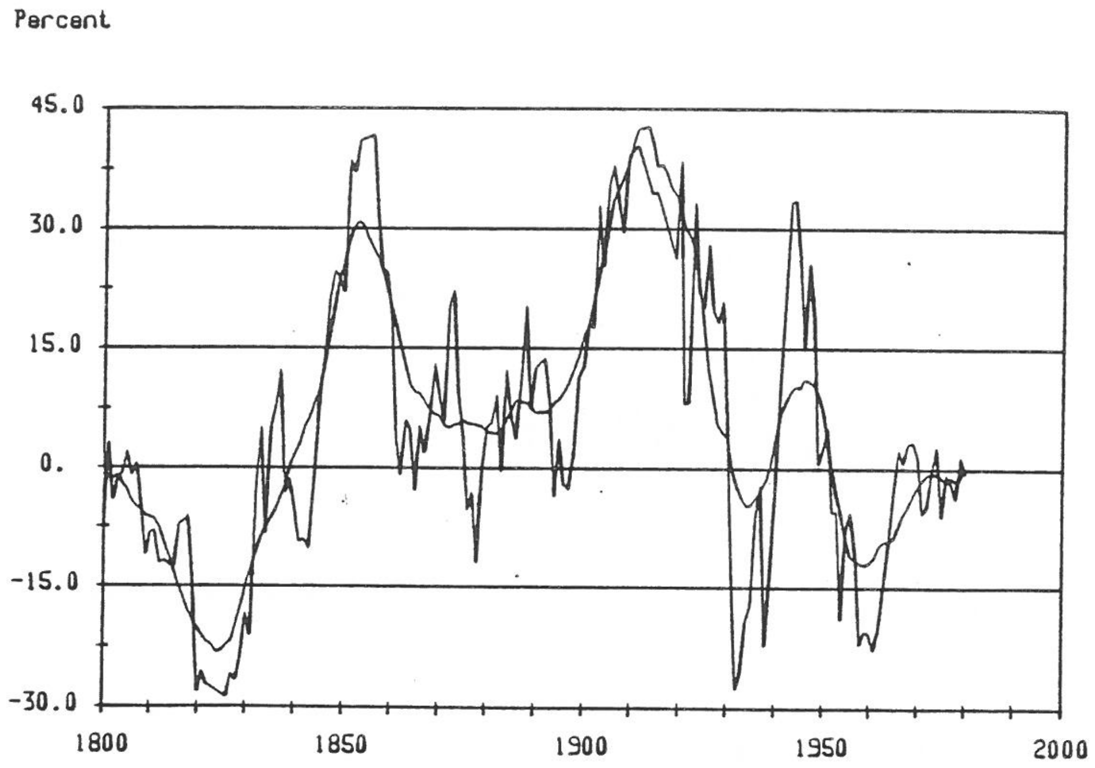


Figure 4.25 Long Wave in Coal Consumption, US.

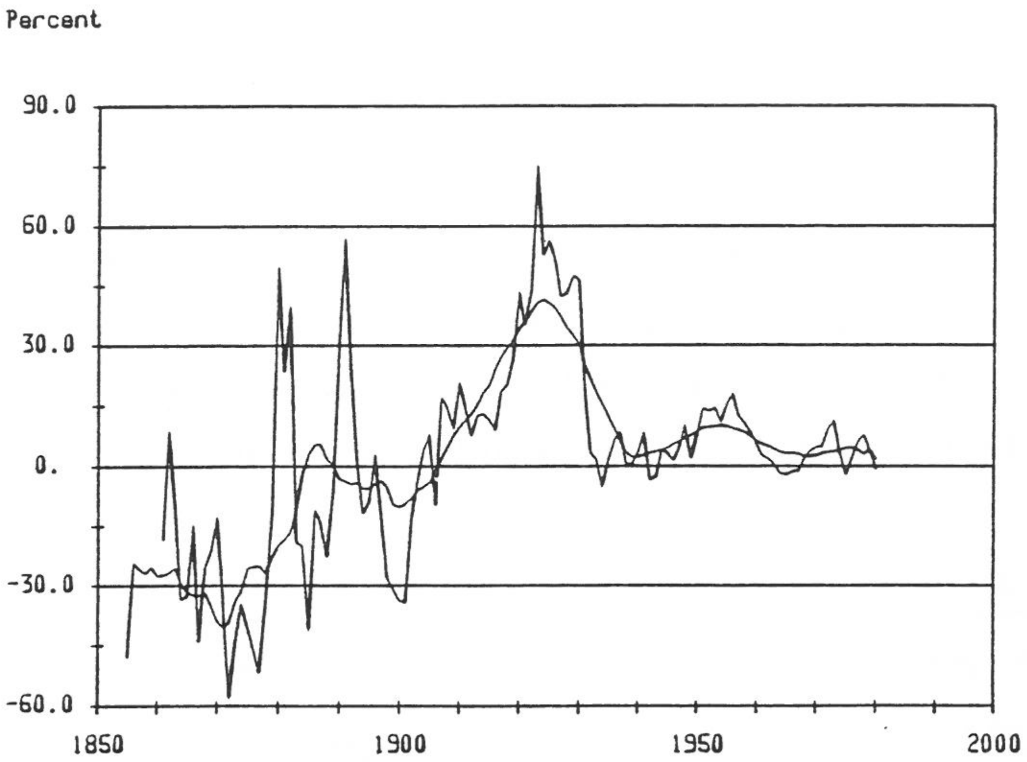


Figure 4.26 Long Wave in Oil Consumption, US.

4.5 Efficiency of Energy Use

There are many ways of determining the efficiency of energy use. The most obvious indicators are the efficiencies of primary energy conversion to secondary and final energy forms. Another possibility is to estimate the efficiency of energy end-use. Examples include the amount of fuel needed for travel, or for space conditioning. All of these efficiences have improved radically since the beginning of the industrial revolution along with the introduction of more efficient technologies. In some cases the improvements span almost an order of magnitude. For example, in 1920 the average efficiency of natural gas power plants in the United States was nine percent, whereas today the best gas turbine power plants can operate with efficiences of almost sixty percent. This improvement spans a period of about fifty years. Over longer periods of time the improvements were even more impressive. Figure 4.27 shows the efficiency improvements of prime movers since 1700 and of lamps and ammonia production since 1800. In Figure 4.27 ε represents the second law efficiency and the data are plotted as the linear transformation of the logistic function on a logarithmic scale, i.e., as $\varepsilon/(1-\varepsilon)$. The efficiency of best current practice improved by almost two orders of magnitude during the last 280 years in the case of prime movers, by almost three orders of magnitude during the last hundred years in the case of lamps and by almost two orders of magnitude during the last eighty years in case of ammonia production.

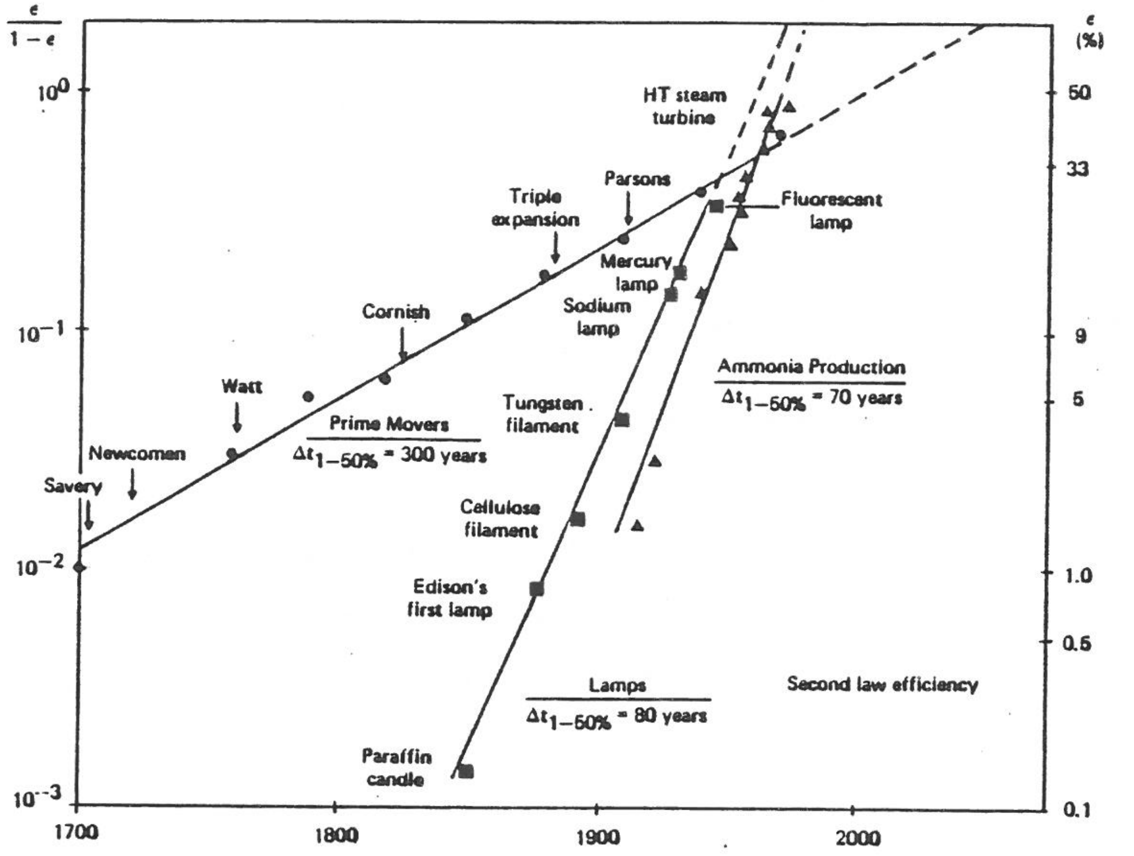


Figure 4.27 Efficiency Improvements Since 1700 (Marchetti, 1979).

All of these efficiency improvements of individual technologies are translated into more effective use of energy and other materials at the level of the overall economic activity. Some efficiency increases result from improved technologies and others from substitution of the old by new technologies. The extent of these changes and improvements can be expressed at an aggregate level by the amount of primary energy consumed per unit of gross national product in a given year. Figure 4.28 and Figure 4.29 show the ratio of energy consumption over gross national product (energy intensity) for the United Kingdom and the United States. In the United States the average reduction in energy consumed to generate one dollar of gross national product was about 0.9 percent per year during the last 180 years. The ratio decreased from more than ten kilowatt-years per dollar in 1800 to slightly more than two kilowatt-years per dollar in 1982. Thus, a regular decline in energy intensity of the whole economy prevailed over a long historical period indicating that energy conservation is a historical process that was discovered as a concept only during the last decade.

In the United Kingdom the energy intensity at the aggregate level of the whole economy portrays a more complex development than in the United States. Initially, energy intensity increased up to the 1870s and then proceeded to decline also at an average rate of 0.9 percent per year between 1900 and 1982. This indicates an interesting parallel development in aggregate energy use in the two countries after 1900 and also poses the question why the energy intensity decreased in the United States during the nineteenth century and increased in the United Kingdom. Most likely this difference is due to the fact that non-commercial energy use is not accounted for in the primary energy consumption for the United Kingdom. If the fuelwood consumption is excluded from the ratio of energy over gross national product in the United States, the same secular trend as in the United Kingdom is obtained: energy intensity increases up to 1920

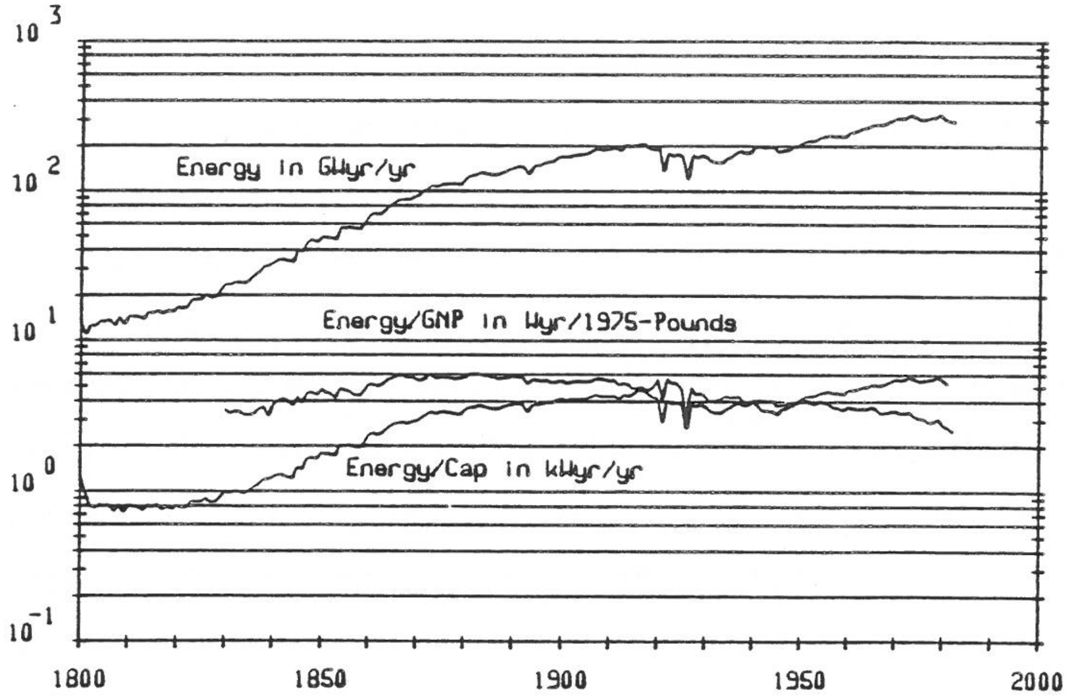


Figure 4.28 Primary Energy, Gross National Product and Energy Intensity, UK.

and than proceeds to decline at a rate of about 0.9 percent per year. The lag of 20 years in reduction of fossil energy use per unit value of gross national product between the two countries is due to the fact that fuelwood was the primary source of energy in the United States throughout the nineteenth century. Therefore, the initial increase in energy intensity in the United Kingdom appears to be due to the substitution of non-commercial by fossil energy sources and not by the actual increase of energy consumption per pound sterling of gross national product. If this was actually the case, then the reconstruction of the ratio of energy consumption over gross national product at an annual 0.9 percent decrease throughout the nineteenth century would offer a method for estimating non-commercial energy use in the United Kingdom.

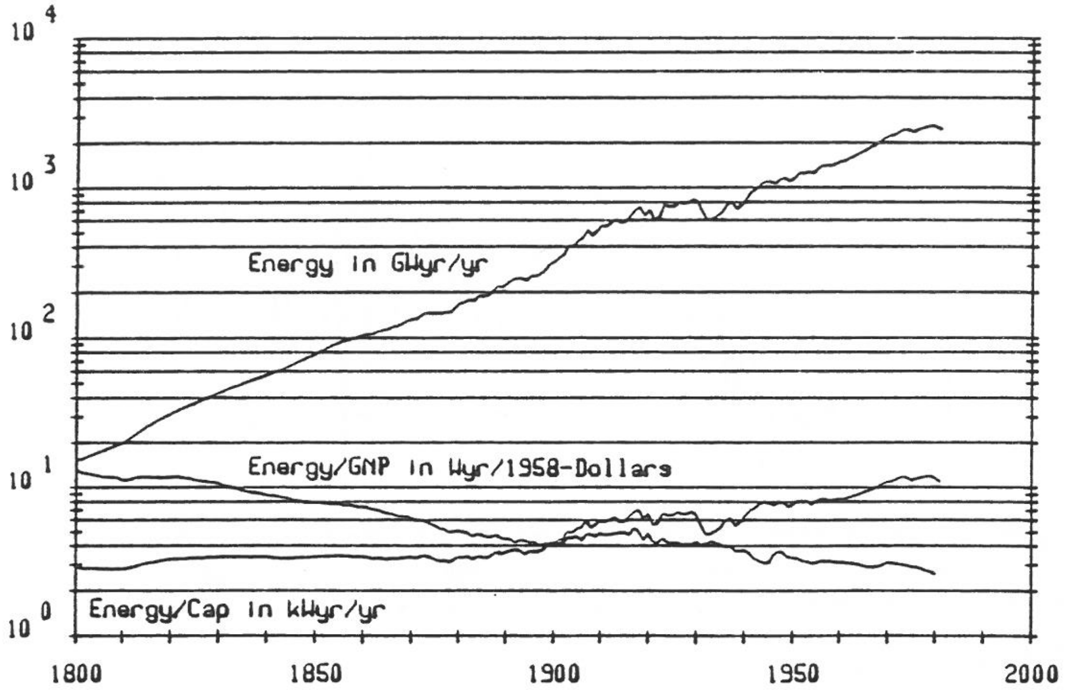


Figure 4.29 Primary Energy, Gross National Product and Energy Intensity, US.

Figure 4.30 and Figure 4.31 show the fluctuations in energy intensity in the United Kingdom and the United States after the elimination of the secular trend by a fifty-year geometric moving average. In both countries the fluctuations show pronounced long wave movements. The long wave in energy intensity also portrays a high degree of synchronization with the price swings since the mid-nineteenth century. During the downswings in prices the energy intensity of the economy decreased more rapidly and during the upswings less rapidly. This means that during the downswing in economic activity general rationalization measures of individual enterprises cause larger energy savings compared with the average historical reductions. As the competition intensifies during the recession and depression, energy savings become an important factor in cost reduction. According to Schumpeter the strive to improve profitability in depressed markets also leads to acceptance of innovative entrepreneurial activities and new

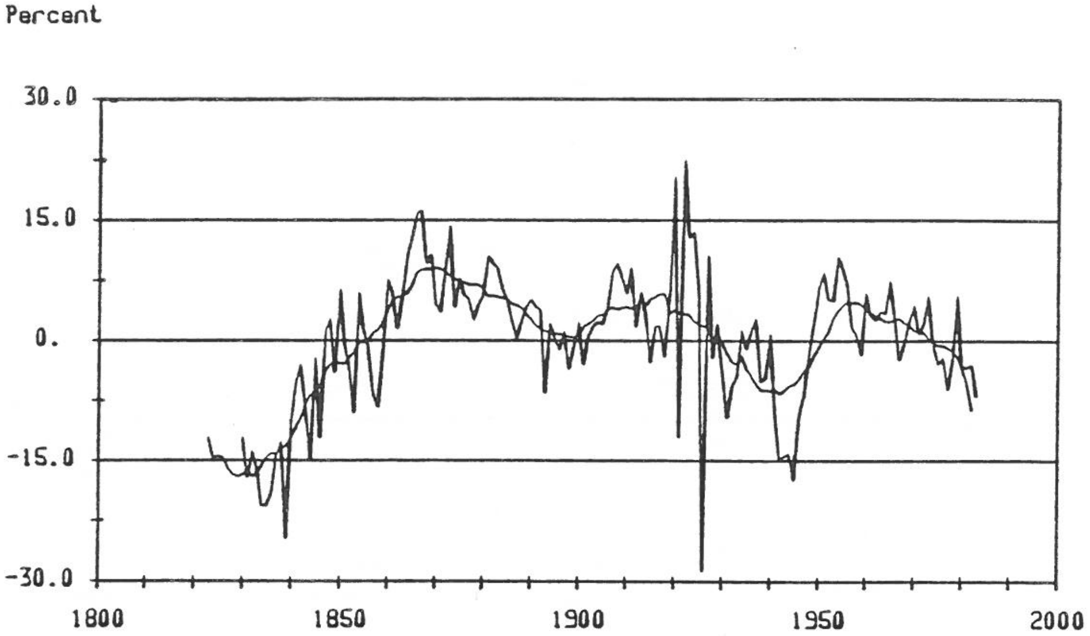


Figure 4.30 Long Wave in Energy Intensity, UK.

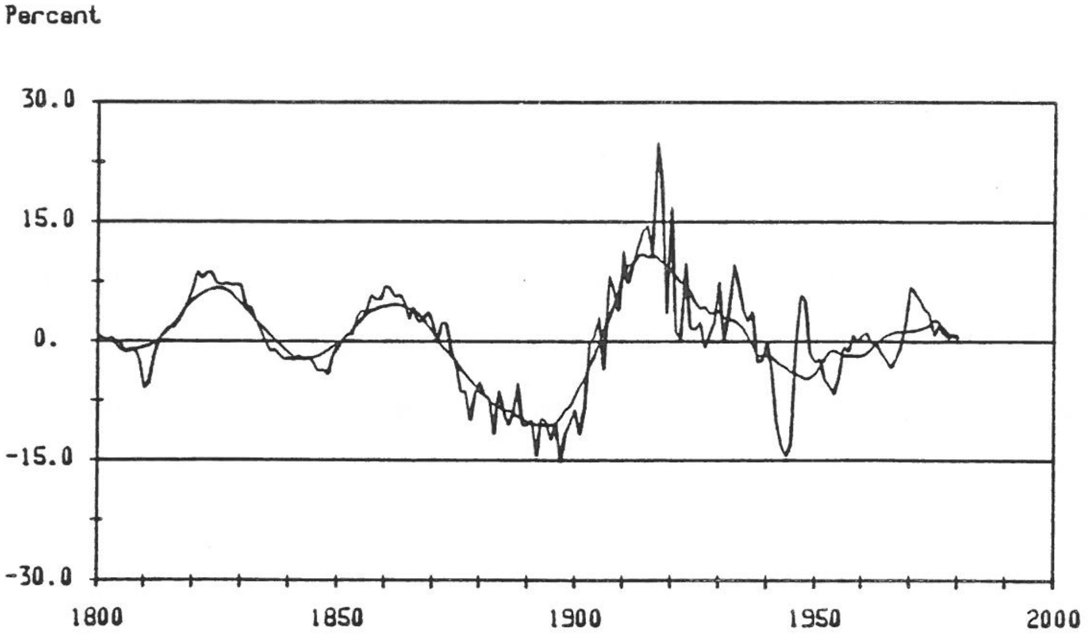


Figure 4.31 Long Wave in Energy Intensity, US.

technologies in spite of the the higher risks involved in such ventures. In time these new practices also acquire substantial market shares so that the relative energy intensity of the economy is lowest just before recovery is initiated. With recovery, new demands and prospects of continued economic growth release many pressures associated with saturating markets, relatively high input costs and low profits. Most of the entrepreneurs in the new growth sectors must intensify their activities in order to meet new demands, and low energy intensity ceases to be an important competitive criterion. New technologies and energy forms offer possibilities for continued expansion in new markets so that relative energy use intensifies. Toward the end of the prosperity period the growth process encounters limits once more. These are reflected in saturating demand and general price inflation illustrated by the long wave of price index movements (see Figures 4.3 and 4.4). Thus, during the upswing energy availability appears not to be a limiting factor in overall economic activity, while during the downswing energy use reductions become important. These reductions are not only due to efforts to cut costs in view of higher energy prices and saturating demand, but also due to a host of social constraints. Many energy technologies, along with other economic activities, become socially and environmentally unacceptable toward the end of prosperity. This means that some diseconomies that were socially acceptable during the growth phase become internalized as additional economic costs or as explicit limits to further expansion. These causes of additional costs appear to off-set the benefits of the economies of scale achieved during the expansion phase. In fact, with the demand reductions during the downswing the large capacities that offered economies of scale become sources of additional costs as excess capacity.

A possible objection to the use of the energy over gross national product ratio as an indicator of the long wave is that its dimension is a composite of energy and monetary units. Consequently, the long wave observed in prices could be reflected in this indicator as the main cause of the fluctuations. This is of course possible, although the gross national product time series is expressed in constant values so that most of the inflationary and deflationary movements should not be an overwhelming aspect of its long wave fluctuations. Figure 4.32 and Figure 4.33 show the long wave in price indices and primary energy intensity and consumption superimposed for the two countries. The fluctuations of energy and prices appear to be synchronous, so that a possible reflection of price movements in gross national product fluctuations would also be synchronized with energy fluctuations. Thus, since all fluctuations seem to be in phase, the energy over gross national product series (energy intensity) indicates changes in relative fluctuations and should not be the artifact of price movements.

The relationship between primary energy consumption patterns and the long wave appears to extend beyond the parallel changes in the relative level of energy consumption and energy intensity with the fluctuations of other long wave indicators. In the case of the United States, we can investigate this relationship between the long wave and the dynamics of primary energy substitution in greater detail because of the availability of estimates for non-commercial energy use. Figure 2.9 showed the primary energy substitution of fuelwood, animal feed and commercial energy sources in the United States. Figure 4.34 shows the fluctuations in energy intensity (energy over gross national product from Figure 4.29) together with energy substitution (from Figure 2.9) for the United States. The upper turning points of energy intensity fluctuations correspond to the saturation points of primary energy sources. The upper turning point that occurred in 1860 is related to the saturation in animal feed substitution, the 1915 turning point with the saturation in coal substitution, and the turning point of the 1970s with the saturation of crude oil. In addition, new energy sources reached

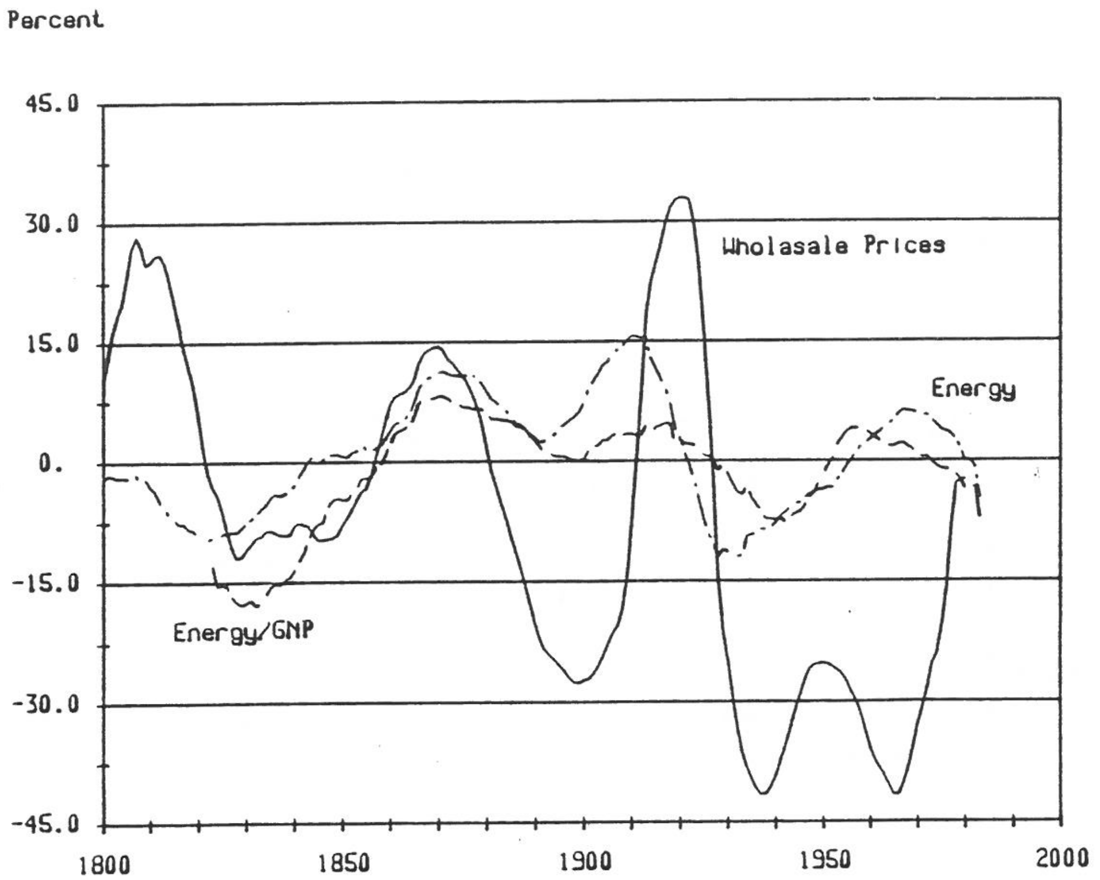


Figure 4.32 Long Wave in Energy Intensity, Primary Energy and Price Index, UK.

one-percent market shares during the times of low energy intensity (during the 1880s and the 1950s). Unfortunately, in the case of the United Kingdom it is not possible to determine whether such a correspondence between the dynamics of energy substitution and the long wave can be observed due to the lack of estimates on non-commercial energy use. Before 1900, coal supplied all commercial energy making the investigation impossible. Figure 2.4 showed that crude oil reached a one-percent market share in 1910, or about a decade later than the lower turning point of energy intensity. Thus, this appears to be a counter example compared with the emergence of new energy sources during the periods of lowest relative energy intensity in the United States. However, Figure

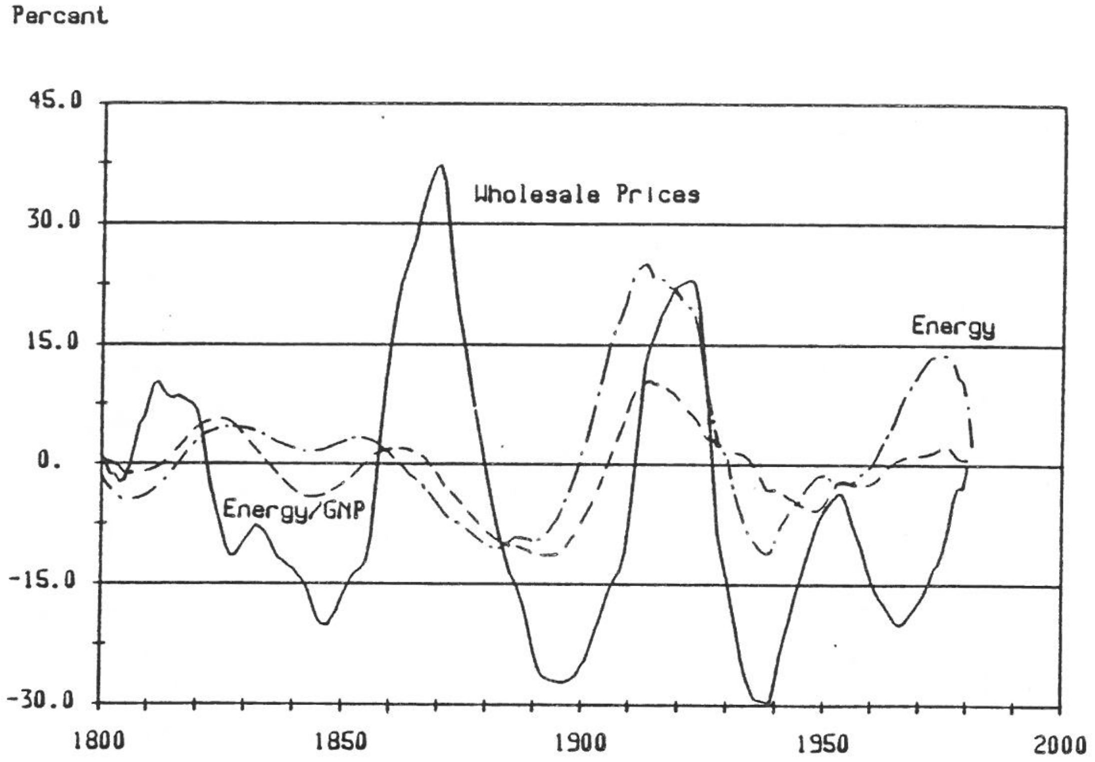


Figure 4.33 Long Wave in Energy Intensity, Primary Energy and Price Index, US.

2.4 also showed that crude oil reached almost a one-percent share in 1900 in the United Kingdom and remained at that level until 1910 when it actually crossed the one-percent threshold. The 1900s, as an introduction time for crude oil, would correspond to the lowest relative energy intensity and lower turning point of other indicators. Thus, the dynamics of energy substitution in the United States indicate a close relation to the time-table of the long wave but it is still an open question whether a similar relationship can be confirmed for other countries.

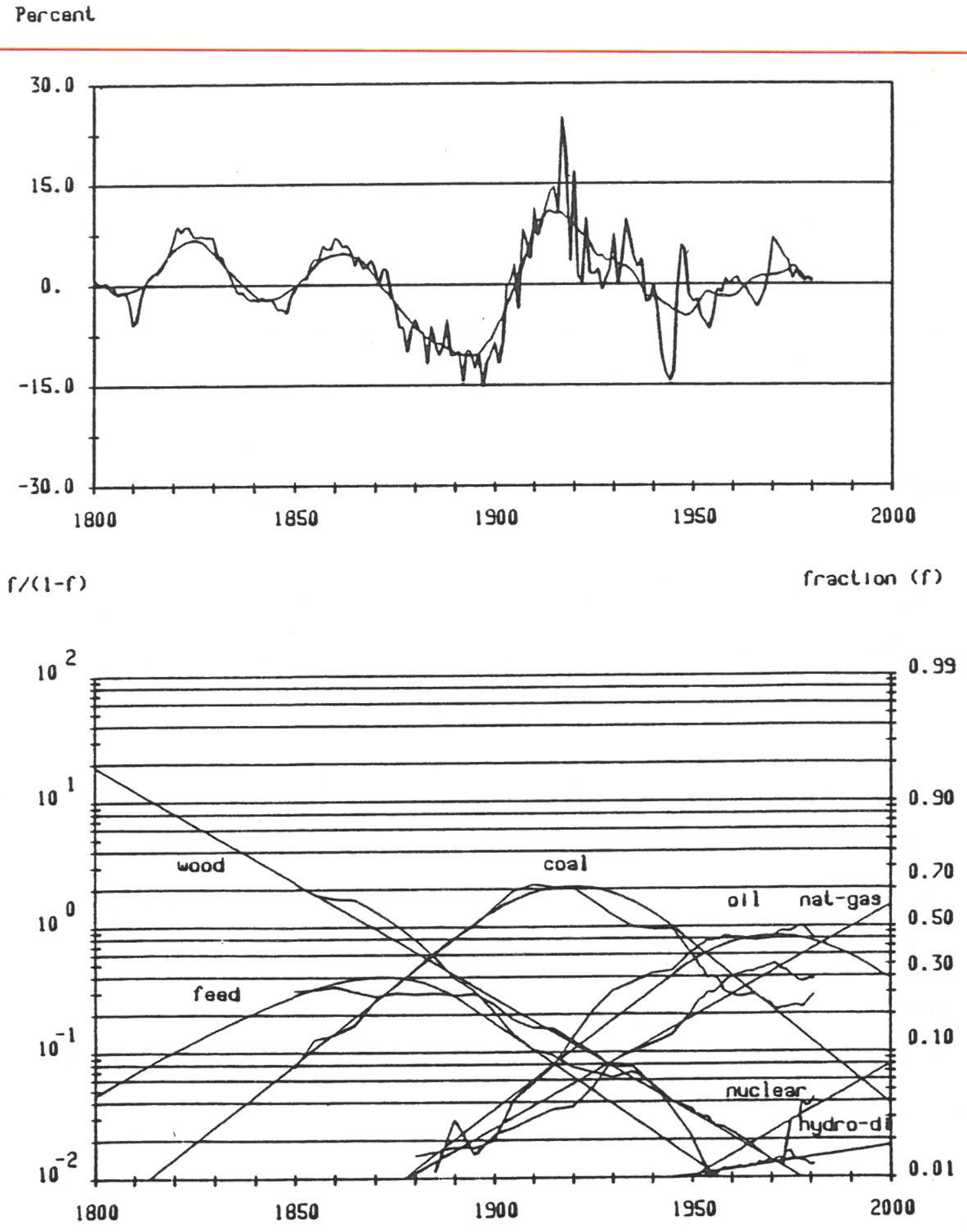


Figure 4.34 Primary Energy Substitution and Long Wave in Energy Intensity, US.

4.6 Physical Indicators: Steel and Ships

In addition to primary energy substitution, we have shown the examples of technological substitution in steel production and merchant ships. Now we will consider these two examples again in the context of the long wave. These two examples have the disadvantage that the time series are relatively short encompassing only the last two long wave fluctuations. This is unfortunately the case with many other possible indicators that are expressed in physical units. Over longer time periods the meaning of the unit measured often changes. During the last hundred years, however, merchant fleets fulfilled similar roles in transport of goods and people although their importance has diminished since the Second World War. Similarly, the quality and mix of steel changed with the introduction of the Bessemer process, but the basic importance and use of steel products has not changed so drastically since 1860 as to eliminate the possibility of considering the data as a continuous series.

Figure 4.35 and Figure 4.36 show the long wave fluctuations in steel production for the United Kingdom and the United States, respectively. Both were derived from total steel production since 1860 in the two countries (given in Figures 2.13 and 2.14) by using a fifty-year, geometric moving average to eliminate the secular trend and an eleven-year moving average to smooth the fluctuations of annual residuals. Two long waves are perceptible in both figures. The amplitudes of the fluctuations are generally higher in the case of the United States than in the United Kingdom. It should be observed that the movements are out of phase with respect to the price swings. The lower and upper turning points precede by about one to two decades the corresponding turning points in prices. This probably means that the markets for steel are more sensitive to the first signs of economic changes and thus respond before other sectors to the

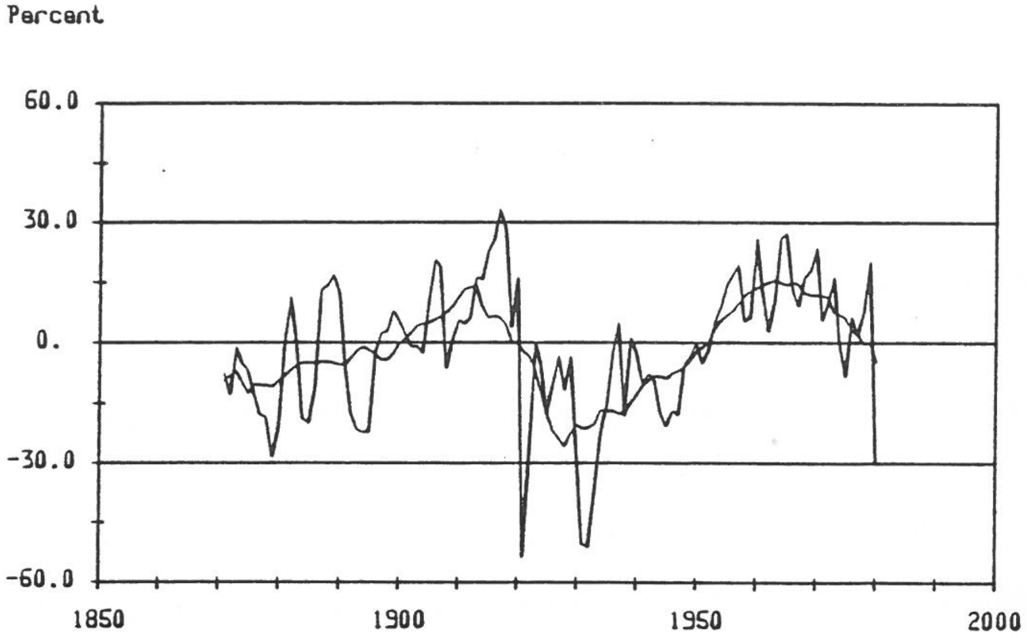


Figure 4.35 Long Wave in Steel Production, UK.

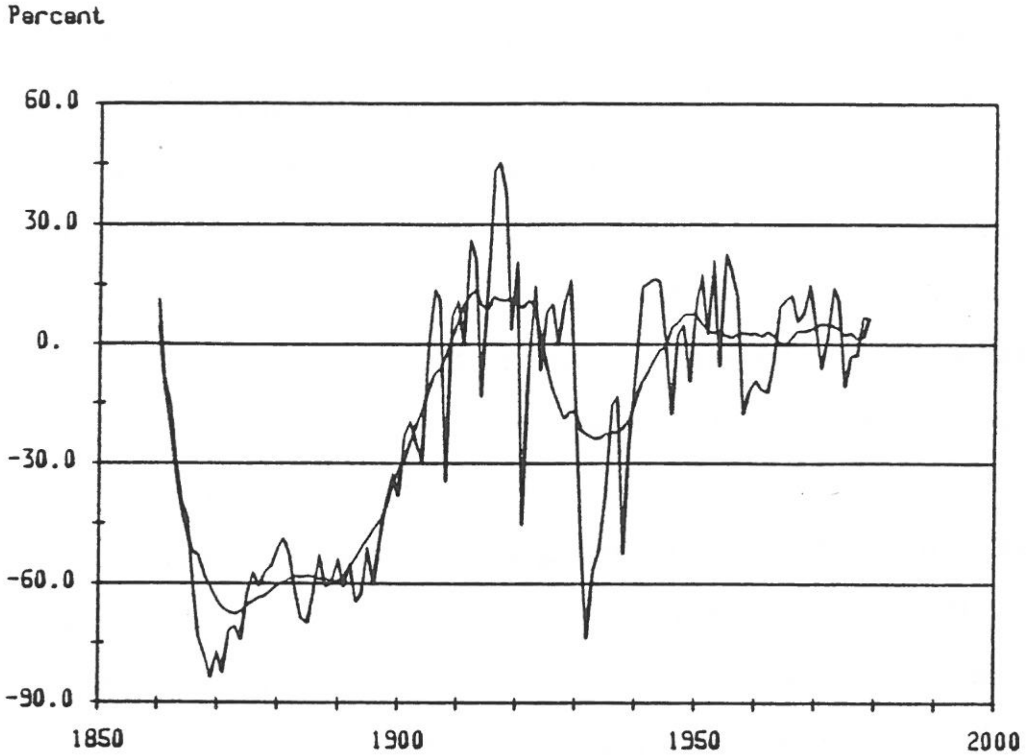


Figure 4.36 Long Wave in Steel Production, US.

emergence of favorable or unfavorable conditions. The reasons for this advanced response of the steel industry may be relatively simple. It is possible that steel, as one of the most important industrial materials, is by and large used in capital intensive goods that have a relatively long life-time and consist of large units. Typical examples from the last century are the railroads and ships, today they are power plants, refineries, large buildings, factories, etc. Even a small decrease in demand for these goods, if it would occur simultaneously, would have an important effect on the reduction of steel production. Thus, it is possible that the first signs of economic change are visible in the fluctuations of steel production because the effect of smaller reductions in many other sectors is amplified when translated into steel demand. If this actually is the case, than one could use the fluctuations of steel production as an early warning for the upcoming turning points of the long wave.

Figure 4.37 and Figure 4.38 show the long wave fluctuations in the tonnage of merchant vessels in the United Kingdom and the United States. The same data were used as in Figures 2.18 and 2.19 where we considered the technological substitution by type of vessel employed by merchant fleets. The data base is again relatively short, especially in the case of the United Kingdom the records extend only to the year 1938, because vessels under five hundred tons are not accounted for in later statistics (see p. 78). In spite of these drawbacks the fluctuations correspond to the long waves in prices in both countries. In the United States a major irregularity occurred after the last wave. A second peak follows immediately after the upswing and downswing between the 1890s and the 1930s. This second peak rises during the 1940s, reaches a maximum in 1950 and than declines during the 1950s and the 1960s. It is interesting to note that this second peak can also be detected in other indicators, but it is not so pronounced as in this case. For example, the fluctuations in primary energy consumption also

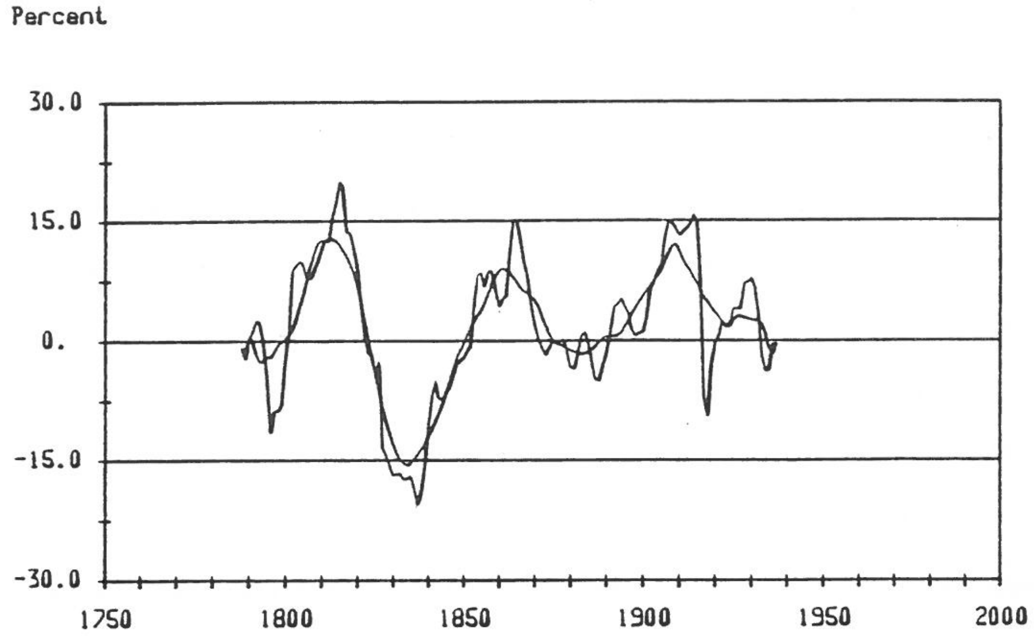


Figure 4.37 Long Wave in the Tonnage of Merchant Vessels, UK.

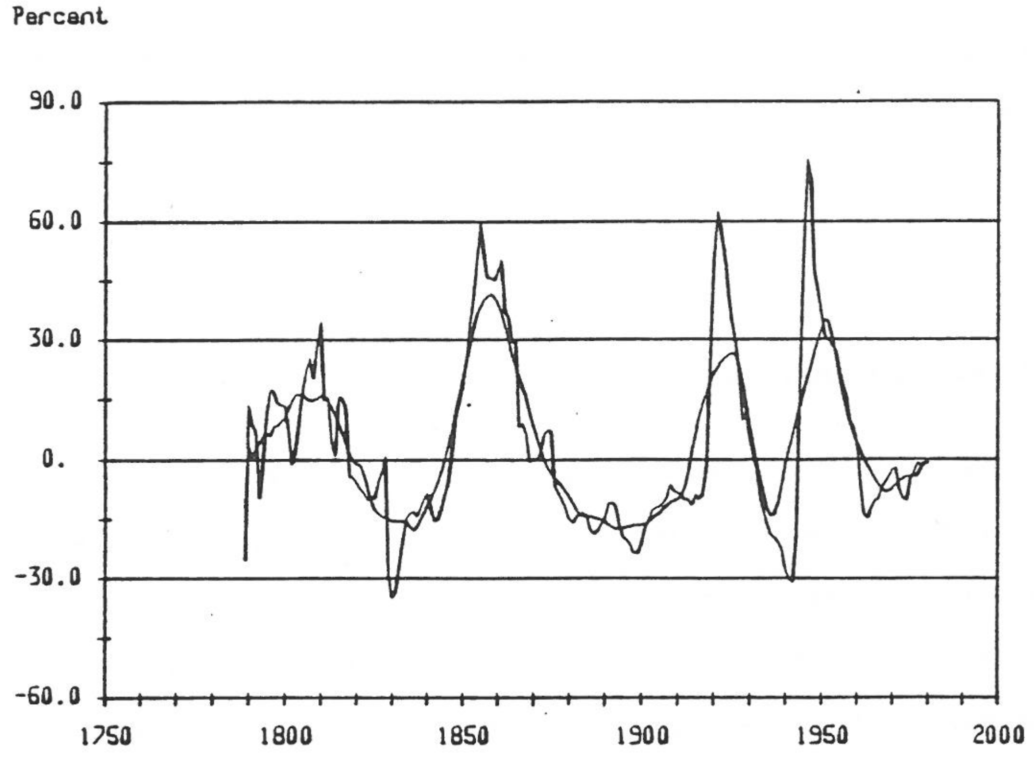


Figure 4.38 Long Wave in the Tonnage of Merchant Vessels, US.

portrayed such a peak during the same period, but it appeared to be only an acceleration and deceleration during the upswing phase of the long wave that was initiated in 1944 with a global peak in the 1980s. Even the wholesale price index shows a subdued fluctuation during the same period with a local peak in 1955, a decline and a renewed rise after 1971. Although, this fluctuation is also present in some other indicators of the long wave, it is by far not so pronounced as in the case of merchant fleet tonnage. Thus, it is not clear from the empirical evidence alone whether the current long wave, which was initiated according to our phenomenological chronology given in Table 4.1, should be divided into two waves of shorter duration, or whether this intermediate fluctuation is an integral part of a single long wave initiated in 1944. If the first alternative hypothesis would be accepted, then the long waves would be subject to an acceleration in frequency because the last fluctuation, as a separate long wave, extends only over three decades. Unfortunately, the time series of the merchant fleet tonnage for the United Kingdom does not cover the period of this last, hypothetical wave so that it is not clear whether the first hypothesis is even applicable to this country, or whether the United States experience is unique.

4.7 Patents and Innovations

A number of theories consider the clustering of innovations as an important cause of the long wave phenomenon. We have already described the work of Schumpeter and the empirical findings of Mensch. One of the major problems connected with this interpretation of the long wave is that there is no obvious and unique way of defining basic innovations which are considered to be the cause or at least very closely connected with the long wave in economic development. The difficulty goes beyond the mere statistical problem of accounting important innovations. The problem is that it is necessary to distinguish between basic entrepreneurial and technological innovations and a much larger number of pseudo-innovations and mere improvements. This selection is usually possible only on a case by case basis because no simple classification criteria exist yet.

Due to these difficulties, the number of patents sealed in a given year is obviously an inappropriate index of innovative activities. In addition, patents are usually granted for an invention that after a period of development can result in an innovation. The period of development can be rather long, sometimes as long as the long wave itself. On the other hand, patent statistics are readily available for long historical periods so that it could be useful to investigate their fluctuations around the secular trend despite all of the drawbacks. Thus, it is probably appropriate to consider patents as an indicator of the intensity of inventions during a given period rather than as a yardstick for innovations.

Figure 4.39 and Figure 4.40 show the number of patents granted each year in the United Kingdom since 1618 and in the United States since 1790. In both countries the number of patents increased exponentially during the nineteenth century. During the eighteenth century the secular trend of patent issues was rather constant in the United Kingdom and the rapid growth during the last century seems to have receded since the 1900s. Also in the United States an asymptotic level appears to have been reached during the first decades of this century.

Figure 4.41 and Figure 4.42 show the fluctuations of the number of patents granted around the secular trend. In the case of the United States, the long wave in patents is in phase with the wholesale price fluctuations. Similar correspondence can be observed in the United Kingdom for the period after the

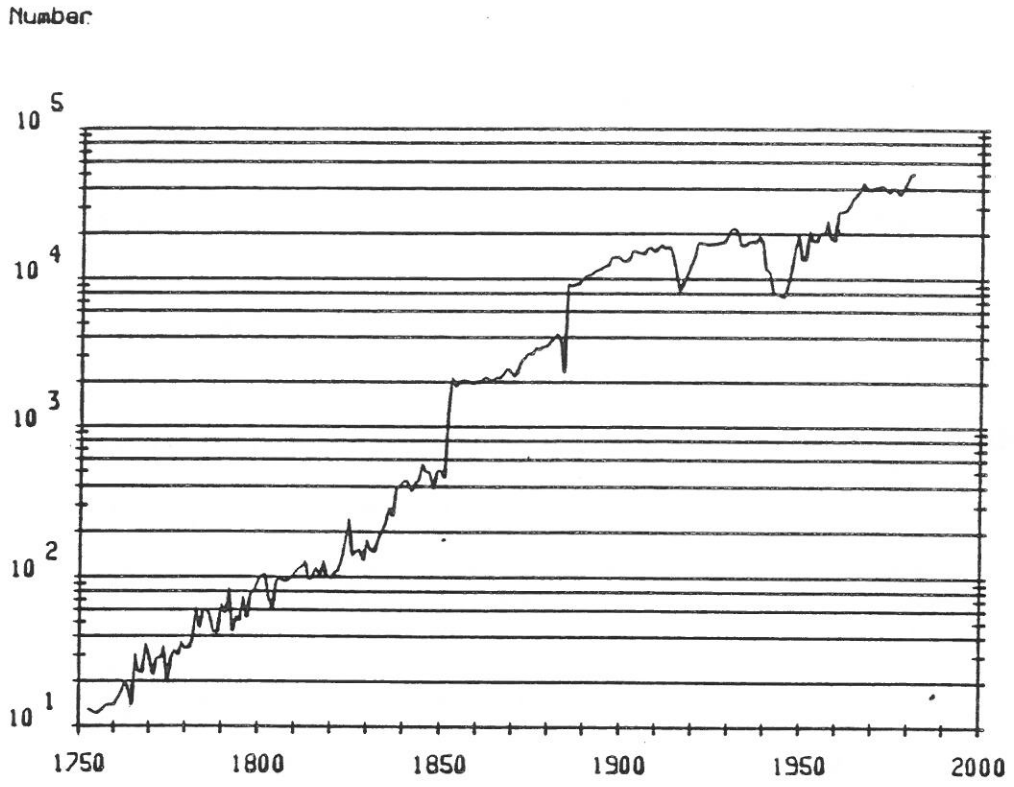


Figure 4.39 Number of Patents Granted, UK.

1830s. During the earlier periods a certain degree of synchronization also exists, but the fluctuations do not portray a pronounced long wave timing. There are two waves of about forty-year duration between the 1660s and the 1740s. From the 1740s to the 1830s, one very long fluctuation can be observed for the patents, so that the long wave behavior appears to be interrupted during a period of about ninety years. The wholesale prices in the United Kingdom display pronounced long wave fluctuations also during this period.

It is intriguing to note that although the patent issues do not necessarily represent a good indication of inventive activities, they display long wave fluctuations over long historical periods that are in tune with the fluctuations of other long wave indicators. According to Schumpeter innovations tend to cluster during recession and the work of Mensch appears to give empirical support to this

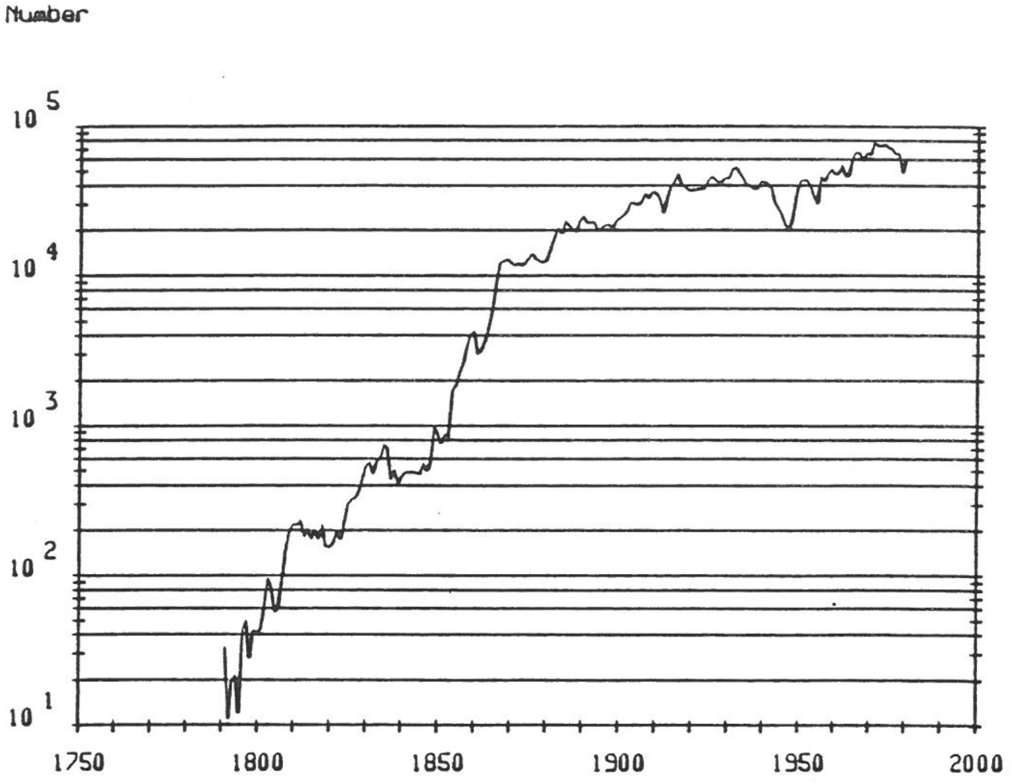


Figure 4.40 Number of Patents Granted, US.

hypothesis. The fluctuations of patents indicate that inventions, represented roughly at this aggregate level by patents, also tend to cluster. However, the clustering of innovations (see, Figure 3.2) are lagged by two to three decades behind the corresponding clusters in patents.

A possible interpretation is that some of the intensified inventive activities lead to basic innovations after a gestation and development period that lasts in the order of two to three decades. Other patents simply represent improvements and are probably applied sooner. This means that toward the end of the prosperity period, the relative number of patents increases in an effort to encounter the problems associated with increasing costs, intensified competition and saturating markets. Most of these patents lead to smaller improvements that probably help the position of individual enterprises, but most of them do not

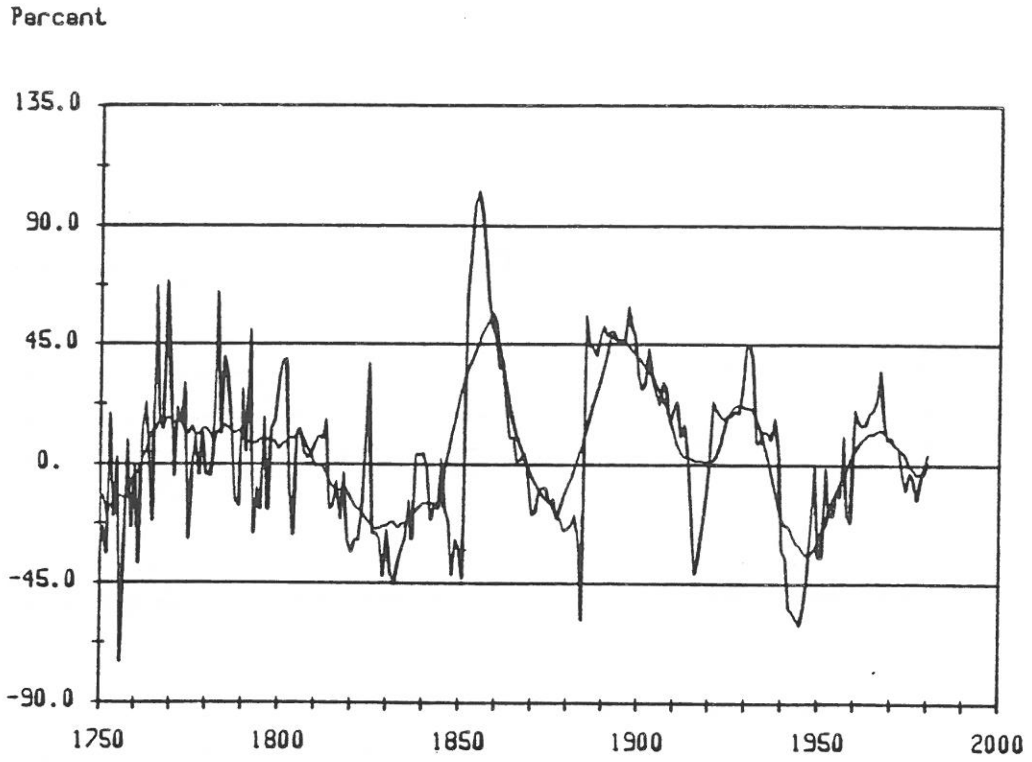


Figure 4.41 Long Wave in the Number of Patents Granted, UK.

represent fundamental inventions that lead to basic innovations and the establishment of new industries. These basic inventions, although in many cases already patented during this period, cannot be applied immediately. During the downswing their development proceeds and applications follow. This also would explain the pronounced clustering of innovations at the beginning of the upswing. Thus, patents including basic inventions tend to cluster at the beginning of the downswing and the clustering of basic innovations marks the end of the depression and the beginning of recovery.

Despite all of the difficulties associated with the definition of basic innovations, van Duijn (1983) compiled a list consisting of 160 innovations that were introduced since 1810. This list has the advantage that it is based on eleven different publications. Van Duijn included only those innovations that occurred in

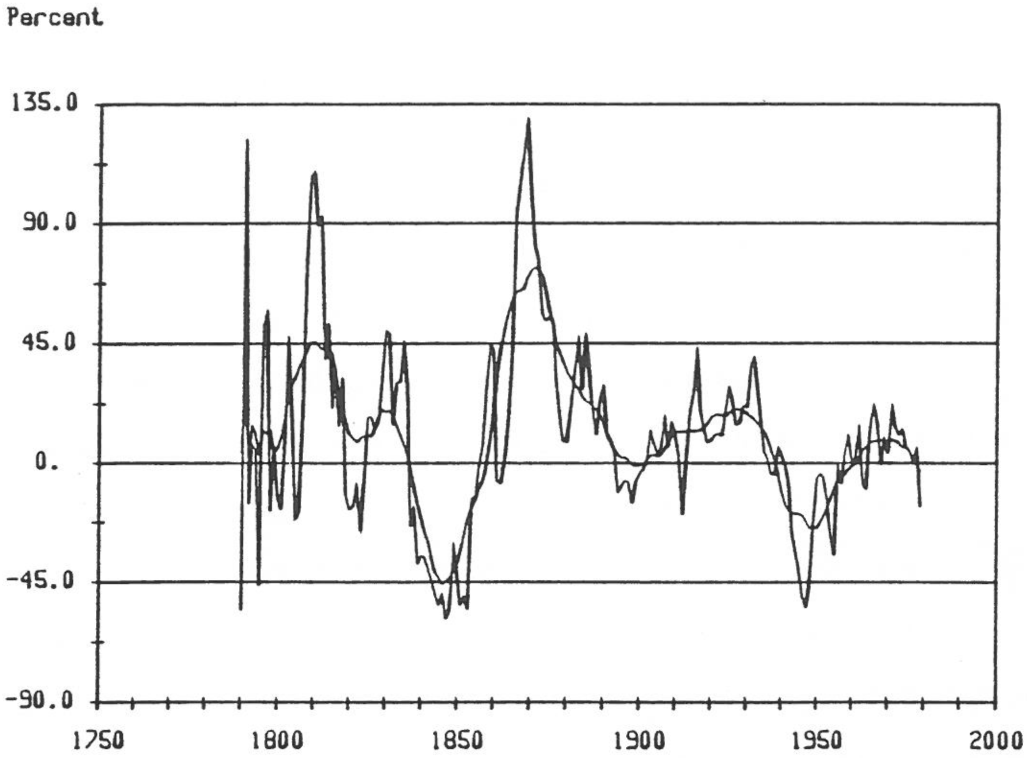


Figure 4.42 Long Wave in the Number of Patents Granted, US.

more than one list giving the assurance that the selection criteria used by one single author do not dominate the record. In addition, the individual lists of innovations that were used were compiled for different purposes than the confirmation of the long-wave bunching effect. "All authors, with the exception of Freeman (but perhaps not the Freeman of that time), have no interest in the long wave and could not be accused of having devised their lists with an intended time pattern in mind" (van Duijn, 1983). There is of course no guarantee that any such list is exhaustive in covering all innovations that could be considered as being basic in the sense that they led to the development of new industries and growth sectors.

We have used van Duijn's list of basic innovations to construct the time series consisting of number of innovations introduced in each year between 1811 and

1971. Figure 4.43 shows the annual fluctuations in the number of basic innovations derived after the elimination of the secular trend by the fifty-year moving average together with the fluctuations smoothed by the eleven-year moving average. The resulting curves indicate pronounced clustering of basic innovations during the downswing periods. This analysis of the basic innovations appears to confirm Mensch's findings. It is, however, curious to note that van Duijn could not find any supportive evidence of the long-wave clustering with the same list of basic innovations. His analysis was based on the one-sample chi-square test at the ten percent significance level for the nine-decade periods between 1870 and 1960 (see van Duijn, 1983). On the other hand, Kleinknecht (1981) found "strong support for the existence of a depression-trigger effect" in the clustering of innovations using exactly the same statistical test. We are confronted here again with the basic problem associated with the long-wave controversy: the long-wave phenomenon appears to be hidden among numerous other fluctuations embedded in the historical statistics. It's confirmation in a particular time series is usually associated with the type of filter used to eliminate the secular trend and other fluctuations. Therefore, it is not *a priori* obvious that the confirmation of the long wave in many different time series is not an artifact of the particular method used and assumptions made by the analyst. We have shown in the case of primary energy consumption in the United States that three different methods of secular trend elimination all lead to pronounced long-wave fluctuations in the residual time series.

Here we can only report that there are time series that do not display long waves apparently irrespective of the method and assumptions used in the analysis. Thus, although we cannot show that the long waves in many time series are not the artifact of the method used in the analysis, at least the converse does not hold: Methods that show long waves in some time series do not necessarily

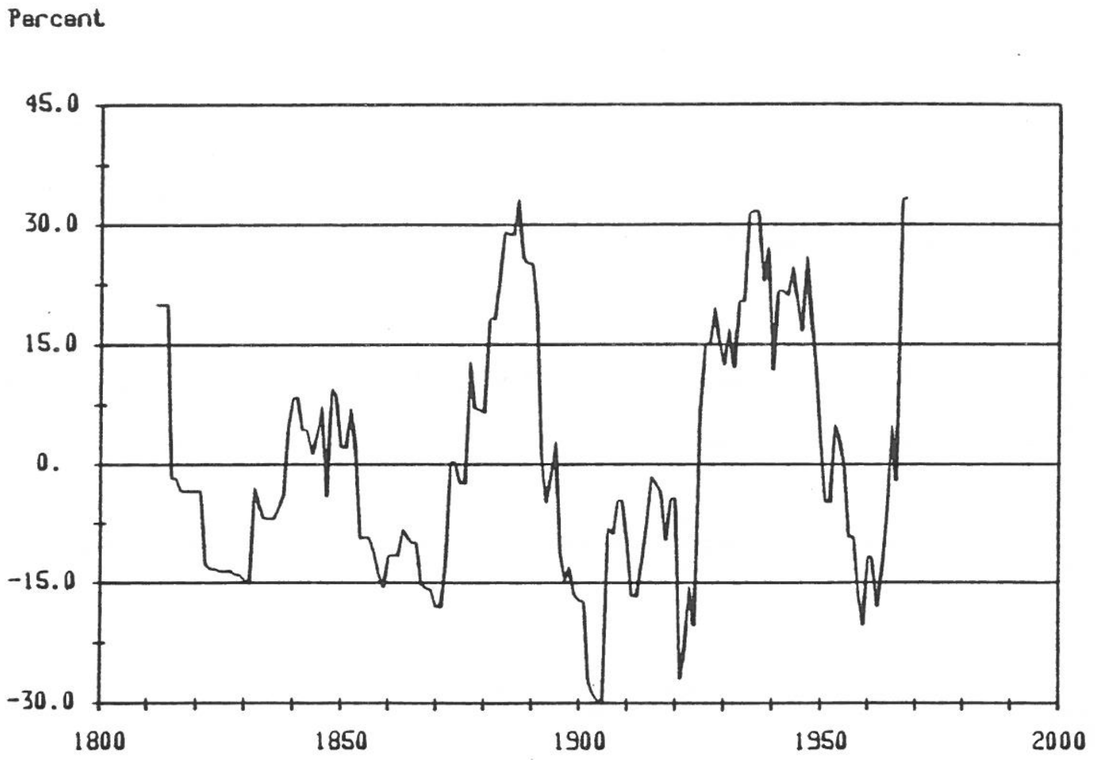


Figure 4.43 Long Wave in the Number of Basic Innovations.

generate long waves in other series. Here it is sufficient to mention that we have encountered a number of long historical statistics that do not appear to contain long-wave fluctuations. For example, the shipments of coal to London from 1665 to 1830 grew rapidly but do not seem to portray any long-wave movements using the same methods for secular trend elimination (see Figure 2.1). Population statistics for the United Kingdom and the United States (given in Figures 4.13 and 4.14) do not show long waves either. In fact Kondratieff found long waves in 25 series and in eleven series he did not detect the presence of long waves (see Garvy, 1943).

Returning again to the relative intensity of innovations, it should be observed that the pronounced peaks shown in Figure 4.43 (based on the list of basic innovations compiled by van Duijn) correspond to the similar peaks in the list of innovations used by Mensch (see Figure 3.2). Figure 4.44 shows the fluctuations of patents in the United States from Figure 4.42 together with the innovation-clusters from Figure 4.43. Here the asynchronous mode of patent and innovation fluctuations is apparent. Innovation clusters are lagged by about two to three decades behind the corresponding peaks in patents. Although the patent fluctuations in the United Kingdom are not shown in Figure 4.44, Figures 4.42 and 4.43 show that the patent fluctuations of the two countries are not perfectly synchronized. In general, there is a shorter lag between innovation-clusters and the patent-peaks in the United States than in the United Kingdom.

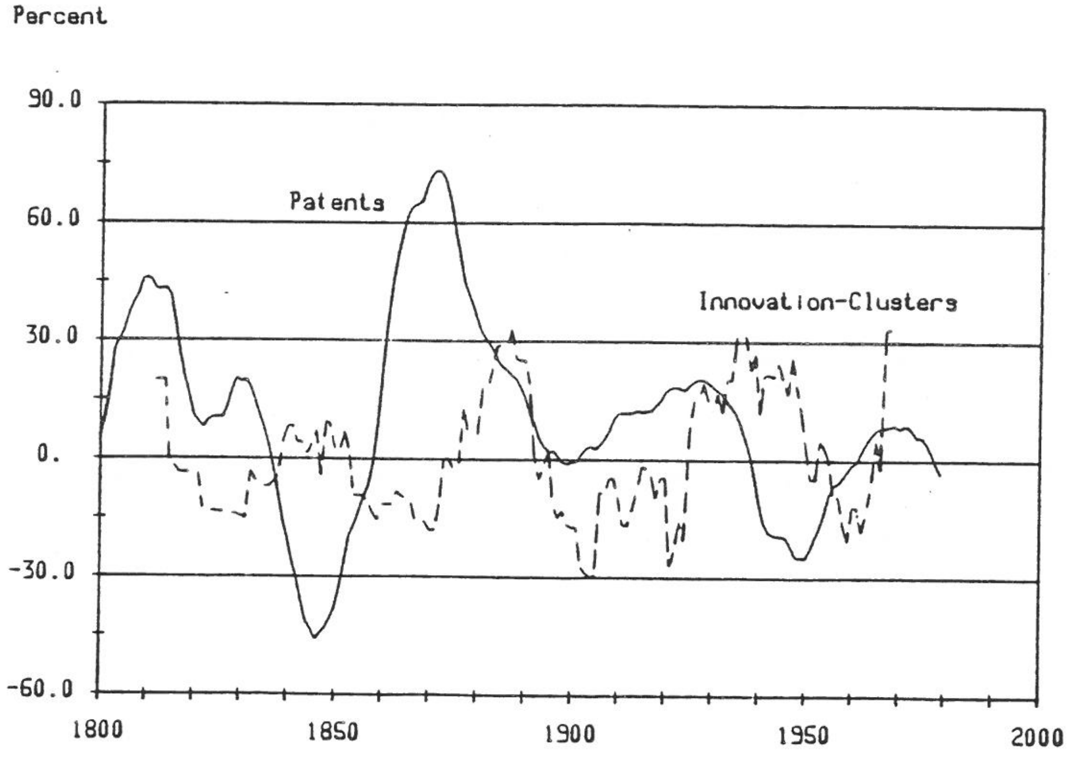


Figure 4.44 Innovation-Clusters and Patent-Fluctuations.

Marchetti (1981) investigated this relationship between basic innovations and inventions that led to the innovations. He used the list compiled by Mensch (1975) and divided the inventions and corresponding innovations into four clusters in accord with the four peaks in basic innovations (see Figures 3.2 and 4.43). Thus, he obtained four lists and plotted the cumulative number of inventions and innovations that occurred by a given year as fractional shares of the total number of inventions and innovations in each cluster. Figure 4.45 reproduces Marchetti's invention and innovation curves. They are illustrative representations of the actual number of inventions and innovations in each cluster obtained by fitting logistic curves to the empirical data. The four invention and four innovation curves are plotted as the linear transform of the logistic function in Figure 4.45 (i.e., as $f/(1-f)$, where f is the fractional share of innovations or inventions in a cluster). Since plotted curves follow linear trends (i.e., logistic paths), they indicate a regular growth pattern within each cluster. If the fifty-percent points (i.e., fractional shares of one-half) on each corresponding pair of invention and innovation curves are connected with each other, the mid-year between these points marks very closely the upper turning point of each long wave. These mid-points are indicated by stars in Figure 4.45.

Figure 4.46 reproduces the invention and innovation curves from Figure 4.45 together with the fluctuations of the wholesale price index in the United States from Figure 4.4 as the indicator of the upper turning points of the long wave corresponding to each pair of invention and innovaton curves. Figure 4.46 shows that more than one-half of all basic inventions in each cluster occur during the upswing of the corresponding long wave and more than one-half of all basic innovations occur during the following downswing. Furthermore, Marchetti observed that the time constants of the invention and innovation curves (i.e., the time it takes until fifty-percent points are reached starting from a one-percent share) tend to get shorter from one cluster to the next. This acceleration phenomenon would imply that innovation clusters should get more pronounced in time. It took in the order of 80 years before 90 percent of basic inventions were

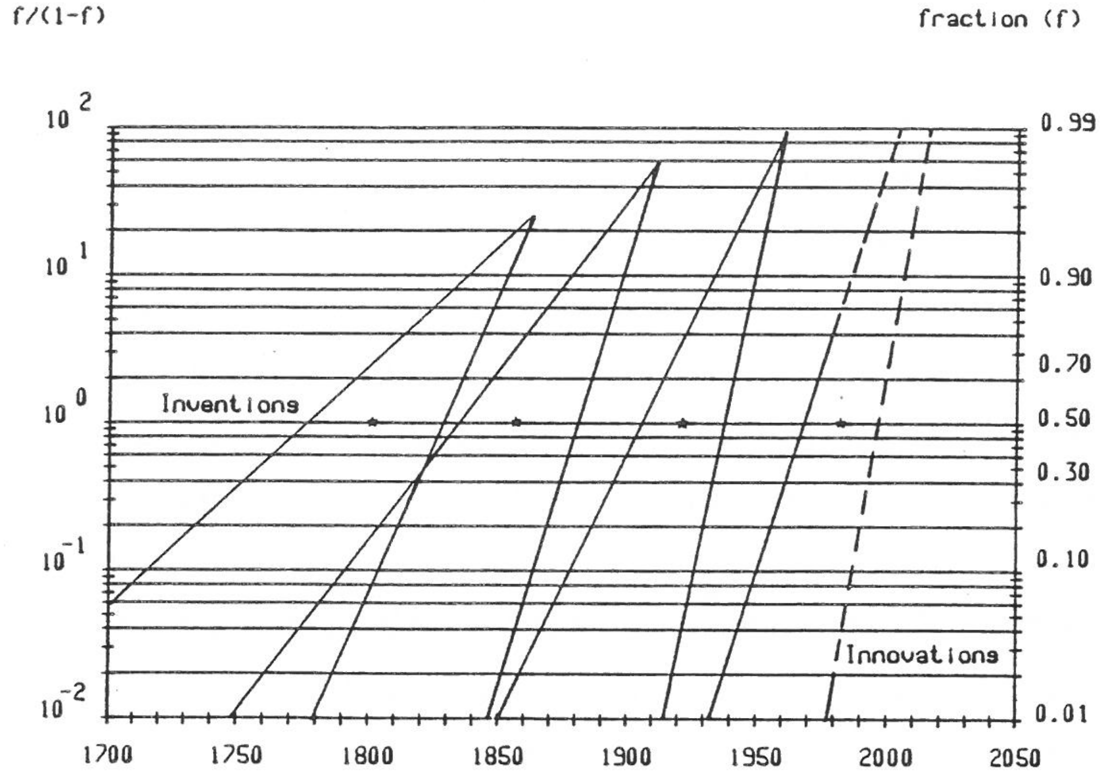
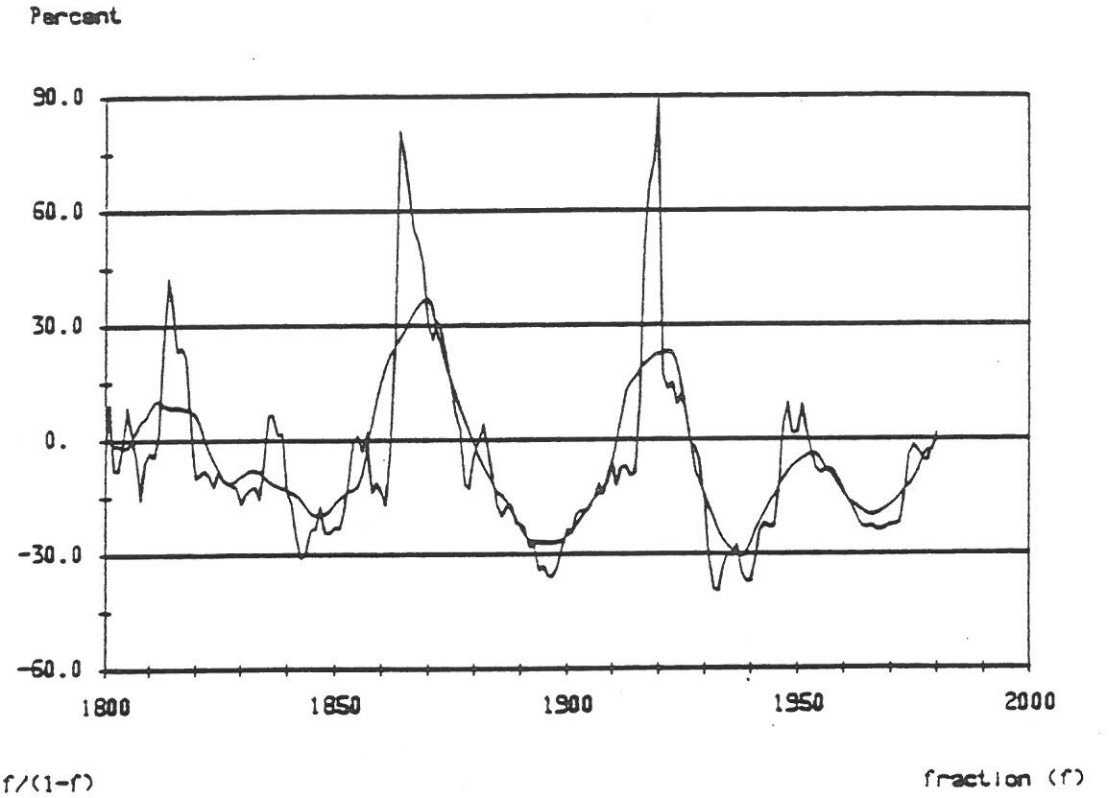


Figure 4.45 Invention and Innovation Pulses, Logistic Fit (Marchetti, 1981).

accumulated that led to the last cluster of basic innovations with the mid-year in 1920 (indicated by a star in Figures 4.45 and 4.46), whereas more than 120 years were necessary for the equivalent accumulation process that led to the previous cluster with the mid-year in 1857. Assuming that the current upswing in innovations is also a sign of the emergence of a new cluster, Marchetti estimated that 90 percent of the necessary inventions would occur by the late 1980s with a mid-year in 1980. This would imply an even shorter time constant for both the basic inventions, of about 70 years, and for the basic innovations. This emerging cluster is indicated in Figures 4.45 and 4.46 by dashed lines. The implication of this result is that we could expect rapid dissemination and introduction of



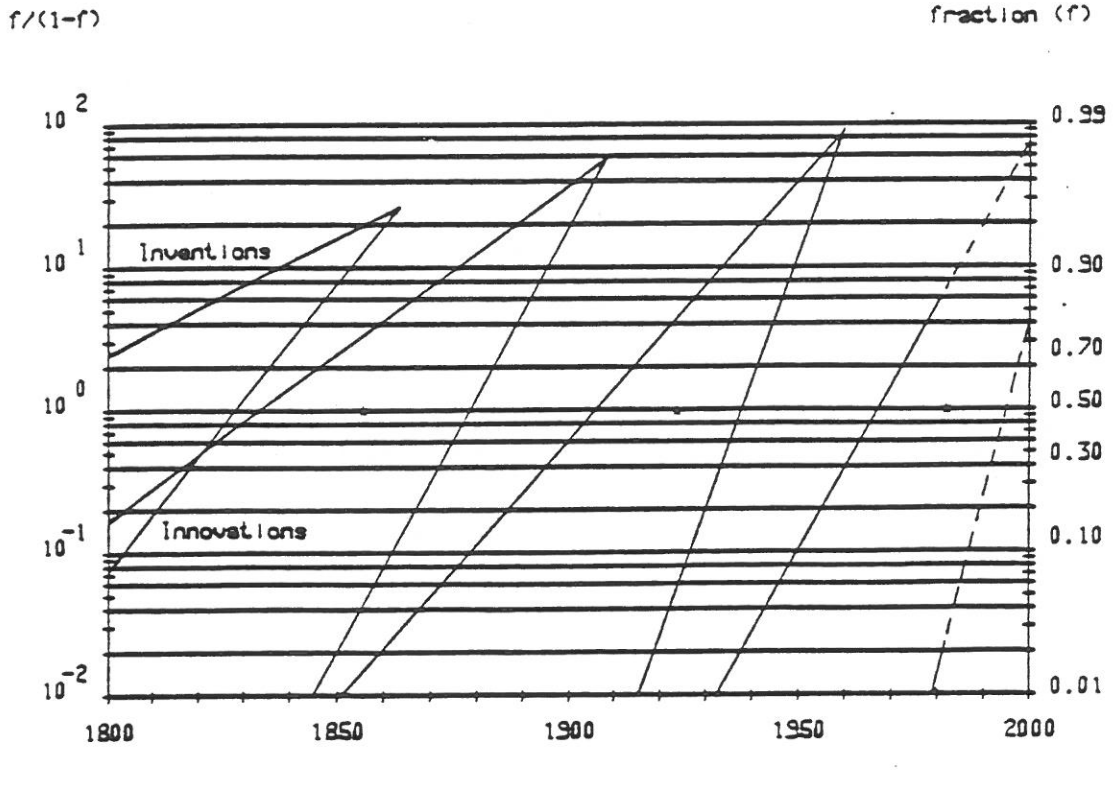


Figure 4.46 Long Waves and Invention and Innovation Pulses.

innovations during the next three decades before the current cluster is completed. Starting in 1980, more than one-half of the basic innovations would be introduced by the beginning of the next century.

5 DYNAMICS OF CHANGE

The phenomenological analysis of many physical and monetary indicators showed a high degree of synchronization in long wave fluctuations. The results are encouraging considering that these indicators are only a rough and approximate realization of the ideal constructs and manifestations of the long wave phenomenon as conceived in various theoretical formulations. The patterns we have distilled out of the empirical indicators represent a kind of "macro-level" description of the long wave phenomenon since these patterns are devoid of the individual historical developments that characterize the evolution of the technological, economic and social development during the last few hundred years. Schumpeter called these individual and unique characteristics "outside factors", not because they are unimportant but rather because they represent ambiguous realization of the historical evolution that consists of recurring patterns at a higher level of abstraction. Considering only individual, unique historical events implies a kind of "symmetry breaking". We have observed recurring symmetry in many patterns after the innumerable factors actually at work in each case were eliminated. At the risk of overgeneralizing, we can state that there is strong evidence that symmetric or at least similar changes in patters of innovations, energy consumption and price *niveau* occur from one long wave to another although the historical content and individual manifestations change profoundly so as to make the symmetry apparent only at the higher level of abstraction. In order to understand the actual mechanisms behind the long wave phenomenon and change in technology, we must acquire better statistical and analytical descriptions of various mechanisms and causal relationships of what we generally call historical experience. This would also imply that we need to understand the course of specific events and their individual manifestations that lead, for example, from a period of rapid growth after the Second World War to the oil shocks of the 1970s, saturating world markets, increasing national debt in many quarters of the world and the economic slow-down of the last decade. For the time-being we can only observe that the particular circumstances change from one long wave to another, but that the sequence of fluctuations and changes at a higher level of abstraction indicate a striking regularity. The annals of business cycles (see for example Thorp and Mitchell, 1926) show that the severe crises or so-called Great Depressions occur regularly during the downswing of the long waves. It suffices here to mention the Great Depressions and financial panics of 1819, 1874 and 1929 in the United States that with small variance occurred throughout the rest of the world. This immediately suggests an obvious historical manifestation of the prolonged periods of stagnation, but this does not answer the question whether these Great Depressions are a necessary characteristic of the downswing. We will try to indicate a possible answer to such questions by considering the regular patterns of technological and energy substitution and innovation clusters. Thus, we will sketch the approximate chronology and timing of pattern-changes during the last three long waves.

5.1 Synchronization and Recurrence

The analysis of technological and energy substitution implies that we consider various technologies or energy sources as mere competitors in an abstract market-place devoid of the particular details of their struggle for market shares, which in retrospect appears only as a historical necessity. In a similar way innovations appear as homogeneous entities that lead to the emergence of new industries or sectors. In summary, we will first consider the energy and technological substitution dynamics and then the correspondence between the chronology of the long wave and the substitution dynamics in the United States. We will not consider the case of the United Kingdom because the quality of the historical material is not as good as for the United States and because the United Kingdom is in many aspects a special case being the source of many changes associated with the industrial revolution that only slowly disseminated throughout the world. The United Kingdom was the world's leading economy practically since the seventeenth century up to the late 1880s, and as such it is not indicative of the structural changes that occurred with considerable lags throughout the rest of the world. Assuming the long waves and dynamics of technology to be a world phenomenon, the United Kingdom represents a leader rather than a typical case during most of the historical period under consideration.

The analysis of technological substitution in steel production, merchant vessels and energy showed that the same basic mechanism can be applied to describe the observed structural changes. In all three cases older technologies were replaced by new ones with regular recurring patterns. In view of the fact that we have also observed strong indications that basic innovations cluster periodically, it is obvious that the introduction of new technologies to eventually replace the old ones is initiated either during or shortly after the clustering of innovations. Before we can consider a possible way of investigating the timing of innovations and subsequent market penetration by new technologies, we will first compare the dynamics of the three cases - steel, ships and energy. Figure 5.1 shows these three substitution cases. Besides the now obvious similarity in the substitution patterns, it should be observed that the timing of the saturation phases is also strikingly synchronized in the three examples. In order to facilitate the comparison we have shifted the curves in time so as to align the

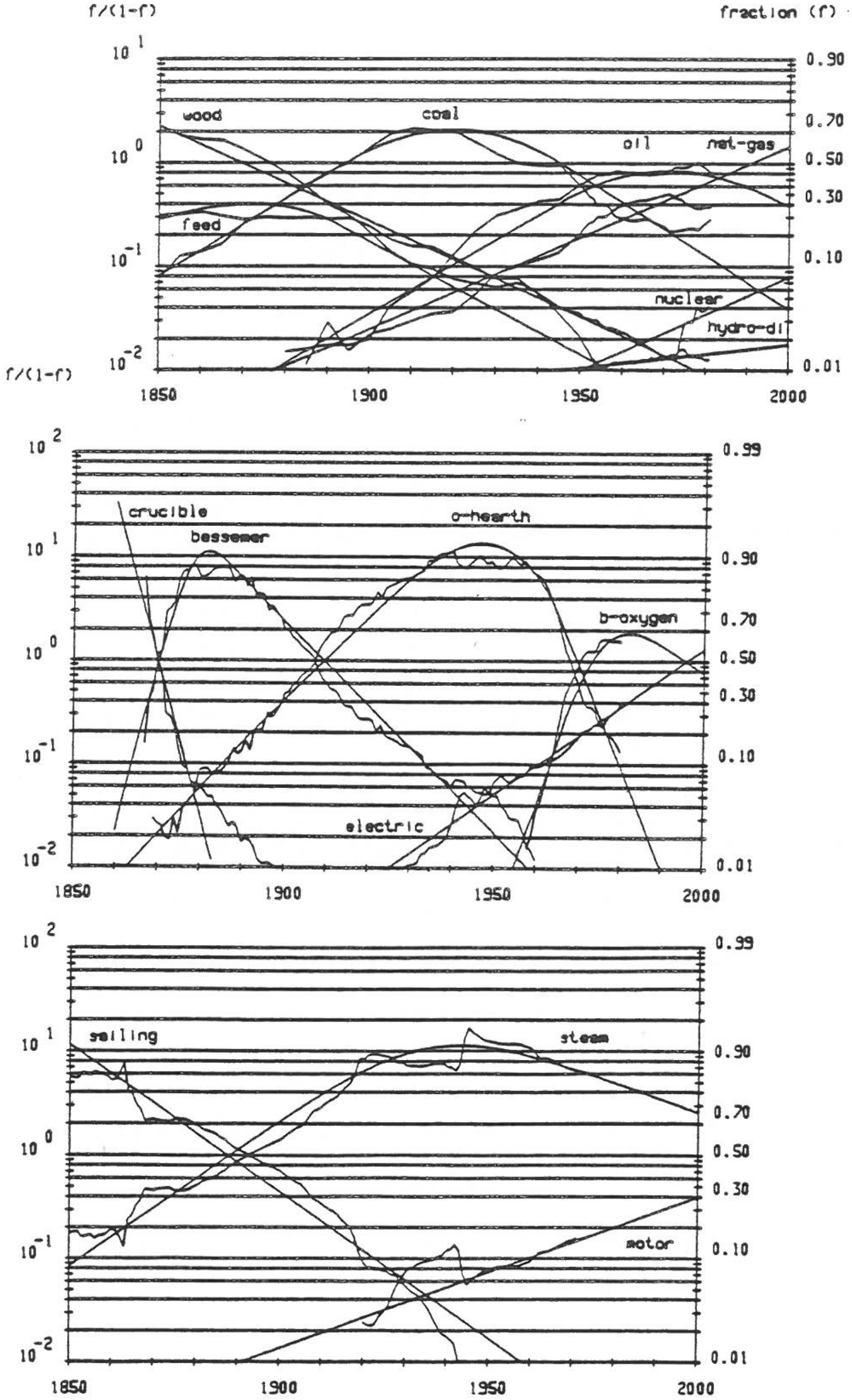


Figure 5.1 Energy, Steel and Merchant Vessels, US.

saturation phases. In comparison to the saturation of coal in the example of primary energy substitution, the saturation of open-hearth steel technology and steam ships is lagged by about twenty years. Once the curves are shifted in time by two decades, as shown in Figure 5.1, other saturation phases correspond to each other as well. For example, the saturation of hay as the energy source for animal feed was reached in the 1870s and the saturation of Bessemer steel about twenty years later. A similar correspondence can be observed for the last saturating technologies - crude oil and basic-oxygen steel. The substitution of other merchant vessels by motor ships corresponds nicely to market penetration of electric steel and to natural gas with a lag of about twenty years. This may be indicative of the continuing synchronization of the dynamic substitution processes in the future, after allowing for the relatively short time lag. It should be observed that the lag of twenty years spans a shorter period of time than the duration of the upswing or downswing phases of the long wave. Although the timing of the introduction of new technologies at the one-percent level differs in the three examples, the change in leadership from the old to the new dominating technology is strikingly similar. The open-hearth steel making process emerged as the dominating technology (in 1907) about 21 years after coal replaced fuelwood as the major source of energy (in 1886). The lag was even shorter in the case of steam ships which overtook sailing ships in 1892. Thus all three takeovers took place within two decades. Half a century later, a similar correspondence can be observed again. Crude oil surpassed coal in 1950 and basic-oxygen steel overtook the open-hearth process in 1969. Again a lag of two decades. Just as in the case of the long wave fluctuations, we find that the substitution dynamics can be characterized by coordinated fifty-year phases of change in market domination from old to new technologies and energy sources.

A possible explanation of this similarity in the substitution patterns is that the specific changes that led to the replacement of old by new technologies and energy sources were interrelated. For example, the new steel processes and naval propulsion systems were dependent on new energy technologies. On the other hand the new energy sources could only be developed with increased intensity of energy use, such as in the new industrial and urban complexes that emerged as the availability of transport possibilities and basic materials increased (symbolized here by steel and merchant vessels). This kind of interdependent lacing of technological development and growth of demand indicates that a certain degree of synchronization in the substitution processes could be expected. Let us make a simplifying assumption and postulate that markets are not created from a vacuum, but that they potentially exist and only have to be taken over by new technologies and energy sources. They represent some kind of niche that can be filled. Since one niche is not usually filled by two identical species, markets also portray a dynamic struggle for dominance. If the new competitor is successful in penetrating a few percent of the market, it usually also wins this fierce struggle for market shares. If viewed from this simplistic perspective, the only prerequisite for substitution of the old by the new are innovations that allow the creation of new technological and energy species. This of course still leaves the question of the precise nature of the fifty-year time constant unanswered. Since we have already shown that the three substitution processes appear to be synchronized after allowing for a two-decade lag in the timing of crucial market saturation and takeover events, we will now consider only the timing of innovations and energy substitution (taking them to be indicative of other technological substitution processes).

Figure 5.2 shows energy substitution, invention and innovation curves (from Figures 2.9 and 4.45) on the lower plot and long wave in energy consumption,

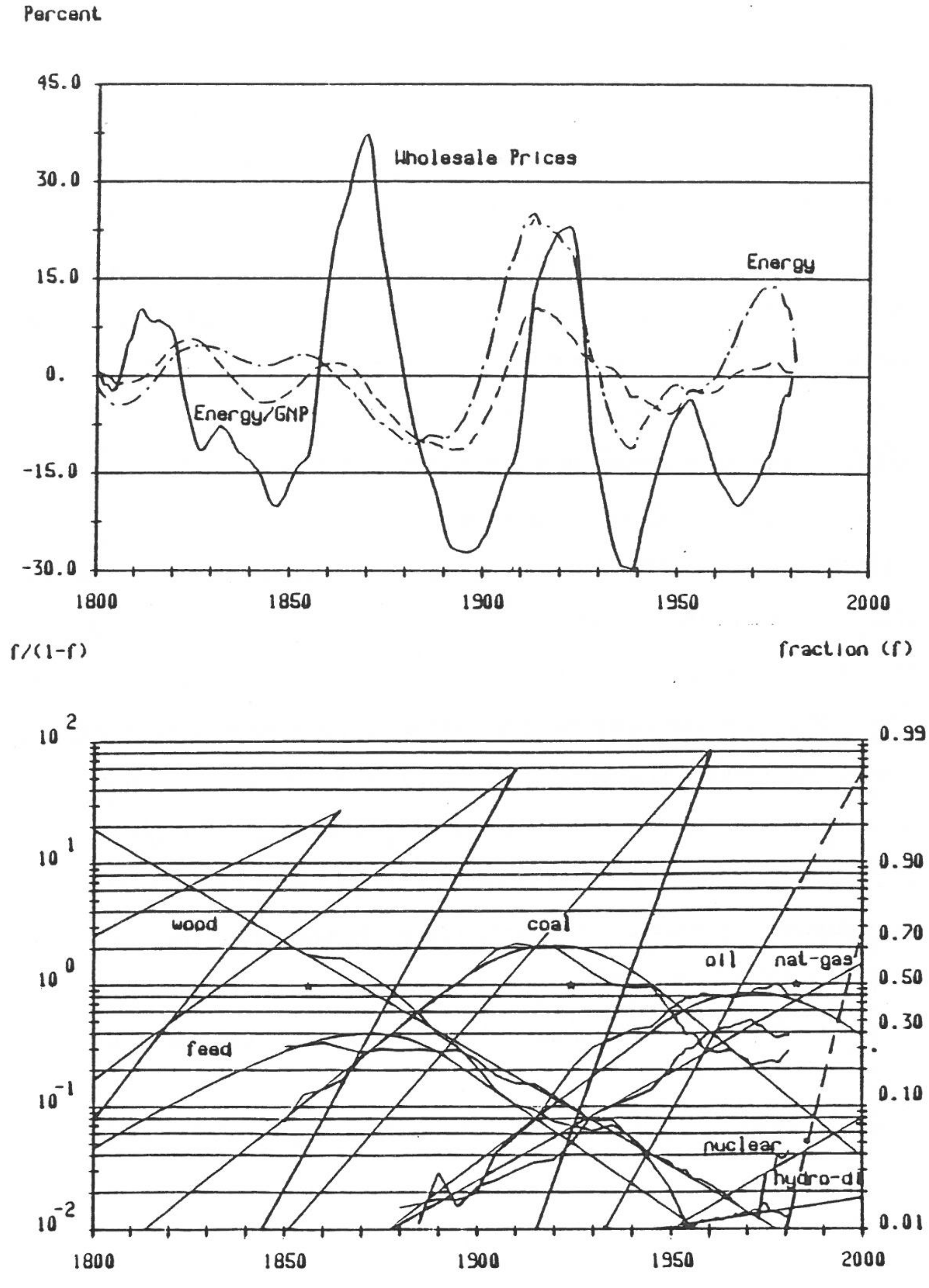


Figure 5.2 Long Waves and Substitution Dynamics.

intensity and prices (from Figures 4.45, 4.33 and 4.4) on the upper plot. *In toto*, Figure 5.2 summerizes the results of the phenomenological analysis of the dynamics of technology and the long wave in economic development. A careful examination of the timing of changes shows that they are all in tune. The saturation periods of energy technologies coincide with the mid-years of invention and innovation curves (denoted here by stars, see also the description of Figure 4.45 above). During the same period we also find the peaks in prices and energy intensity (see also Figure 4.33). Innovation pulses emerge later during the downswing in prices and energy intensity (in the upper plot), or after the onset of senescence of dominating energy sources (in the lower plot). We define here an energy source as dominant in a given market if it controls the largest market share during the given period. In other words, the innovative activity intensifies after the saturation of the dominant energy source has occurred (assumed to also indicate the slightly delayed saturation of other dominant technologies). This period of intense innovative activities continues until the new energy sources (and later other new technologies) have replaced their predecessors from their position of market dominance. This period from saturation to replacement in dominance lasts in the order of 25 years, or about as long as the downswing phase of the long wave which is characterized in Figure 5.2 by the fluctuations of energy consumption, intensity and the price index. By symmetry, the upswing of the long wave starts after the clustering of innovations and is paralleled by the growth of the new energy source from newly acquired dominance to saturation.

6 SUMMARY AND CONCLUSIONS

The fact that all of the events that characterize profound changes in technology and economic structure occur in tune is striking, but it leaves many questions open. For instance, we have observed that technological substitution in steam propulsion and merchant vessels is lagged by about two decades behind the equivalent events in energy substitution. This would imply that these other dominating technologies do not saturate during the end of the prosperity phase, but rather during the onset of the downswing, well into the period of innovation clusters. Perhaps this is an artifact of the choice of technological substitution processes in that they are very closely related to the changes in the structure of the energy system. Yet, given the sparse statistical records, it is difficult to find other examples that span equivalent historical periods. These difficulties are further compounded by definitional problems. It is not obvious, for example, how to define the market place (i.e., the measurement units) for the various information technologies used since the emergence of the industrial age. Similar difficulties arose in other cases we were considering for analysis.

Nevertheless, the importance of the energy system and related infrastructural developments appears to be crucial with respect to the observed pulses in economic activity. For example, the construction of great canals throughout Europe and the United States during the eighteenth and beginning of the nineteenth century was initiated by the ever increasing need to transport timber and other goods in larger quantities over longer distances. Later, railroads caused a similar boom period basically due to the same reasons - the concentration of production in urban areas required a more efficient transport system that also helped in the acquisition of new and larger markets. Thus, canals and railroads expanded existing markets and "created" new ones for new products. In terms of the energy system the large canals are associated with the transport of fuelwood that was still the primary source of energy in many industrial activities such as iron smelting. Draft animals and their energy needs in terms of feed were still required as the major source of shaft power for local transport in addition to direct use of water power and the emerging steam technologies. The railroad era is very closely related to the widespread diffusion of steam and coal related industries.

In terms of our chronology of long wave pulses given in Table 4.1, we will call the upswing phase from 1773 to 1810 the "age of canals" and the upswing from 1840 to 1869 the "age of railroads". Accordingly, we call the upswing from 1895 to 1920 the "age of electricity" because of its significant contribution to the rapid development of new industries and communication technologies. The last upswing, initiated in 1945, we symbolically identify with the motor vehicles, aircraft and petrochemical industries, and according to our chronology of long waves it terminated during the 1970s. Unfortunately, it is not possible to time this last turning point with any precision, but in view of the empirical evidence in the synchronization of technological substitution processes and the pulses in innovations, energy efficiency and other indicators, it probably occurred during the "oil crises" of the early 1970s that mark the saturation of crude oil and its eventual replacement as the dominant source of primary energy. Let us assume for the sake of naming a particular reference year that it in fact occurred in 1973. If this were actually the case then, according to the long wave chronology, the next turning point could be expected sometime around the turn of the century. Going further into the future the following upswing phase could be expected to last until the 2030s.

The overall picture that emerges suggests that each upswing phase is associated with large infrastructural development. This development first opens many new product and factor markets and toward the end of the prosperity phase leads to eventual saturation of these markets and full adoption of the technologies that were introduced during the recovery period. As a result, the increase in innovative activity declines during prosperity following the large productive increases and infrastructural development as growth technologies become fully adopted throughout the economy. This was essentially the process that occurred during the last ten to fifteen years. The innovation pulses from Figure 5.2 indicate that during the next fifteen years about one half of the innovations of the next upswing phase would be introduced. Thus, after a period of almost three decades of relatively low innovative activity, the next three decades would bring another pulse of innovations that are needed for the next upswing phase. Accordingly, most of the basic inventions that could be translated into innovations already exist today, so that the scientific base for the new wave of technologies is, in principle, known. It is no more a question of expecting the unexpected, but rather of identifying the selection principles that will determine which of the innovations are viable and would result in basic innovations. For example, Figure 5.2 also indicates that during the next three decades we can anticipate relative improvements in the energy efficiency of the economy (i.e., reductions in the amount of primary energy consumed per monetary unit of gross national product in real terms). Thus, we can expect further dissemination of energy efficient technologies and institutional measures during the downswing phase until the end of the century. As far as energy technologies are concerned, the market penetration analysis suggests natural gas as the best candidate for eventual dominance as the major energy source during the upswing period after the 1990s. Natural gas is the cleanest fossil fuel and from that perspective alone it is attractive. It also promises well as a very efficient source of electricity and clean fuels. Widespread use of natural gas would require new infrastructures for the long-distance transport, conversion to fuels and electricity, and distribution to the final consumer. Thus, construction of large grids and new industries based on natural gas would be required. Candidates for future growth sectors related to the wider use of natural gas range from technologies for controll and management of large, distributed grids for transport and distribution of energy and other goods, to bio-engineering technologies that would allow for greater efficiency and low-temperature chemical and industrial conversion and production processes based on methane and electricity. Thus, enzymes and microchips may be the hardware that could allow the transition to a methane-based energy system. These are just some of the possible candidates, but they are consistent with the apparent requirements that emerge from the overall pattern of economic pulses and technological substitution dynamics since the beginning of the industrial revolution. Before these and other new technologies could expand during the next upswing, the next three decades would bring a period of renewal and "creative destruction". A period of rapid (relative) deflation can be expected together with prolonged unemployment and an economic slowdown. These are the selection mechanisms that in the past distilled the successful from a wide range of promising technologies and entrepreneurial innovations. The existing patterns will have to be destroyed before new ones can emerge and their destruction will mark the beginning of renewal and a promise of prosperity.

Most of these speculations about the nature and timing of future events is based on the dynamics of equivalent changes in the past. Some of the patterns of these dynamic changes can be projected into the future. Our analysis of the market substitution mechanisms indicated that the invariance of the timing of market saturation and takeover times also allows projections over periods that span the duration of the long wave. Similarly, our analysis of the long swings in many indicators, ranging from energy efficiency to price fluctuations and Marchetti's invention and innovation pulses, provided strong historical evidence that these events are precisely timed and invariant.

Perhaps the most important question is why the clock that tunes such events as the innovation pulses, dynamic changes in technology and long waves in economic activity operates on a fifty-year scale. Since we have shown that the events that mark structural changes are synchronized and follow a logical order, the question of the time-scale invariance is crucial. If it were answered all the other events, since they occur in logical order apparently as required, would fit the grand pattern like pieces of a puzzle. If we use the analogy of synchronization in an electronic computer as indicative of the system in which we are elements, then an obvious answer suggests itself. Computer clocks synchronize individual events in accordance with the reaction time of its elements. Assuming that the active life of a human being spans a period of about fifty years, say from age 18 to 68, then the clock of human systems should also operate within the same period. There are many reasons to assume that inertia and slow absorption do not allow profound changes to occur more often than once per human generation. This is of course just an analogy, but it is nevertheless hard to imagine that people and societies that have invested large resources in creating and adopting to a given environment would accept profound structural changes unless technologies, markets and social institutions had reached the limits imposed by the process of growth or by outside factors. In this analogy we implicitly assume that a human generation behaves in a somewhat predetermined mode - once it accepts a certain pattern of behavior this pattern is not abandoned as long as it functions. Over the time-horizon perceptible to a human being, which is probably not longer than a life-time, human institutions and technologies grow to limits, causing a period of renewal which is manifested by economic depressions and blossoming of innovative activities. New technologies and innovative entrepreneurial activities do not diffuse immediately due to the same reasons - they have to withstand selection processes over long time periods before their viability as a replacement for old methods is accepted. Since we have seen that technologies and markets are interlaced, once we accept a natural rhythm for profound changes to be about fifty years, the synchronization of the pulses and technological changes follow as a direct consequence of interdependencies.

7 POSTSCRIPT

In historical description, we are inevitably influenced by the fact that we observe the recorded phenomena *ex post* with the benefit of hindsight. We have investigated the evolution of technological change and the long wave in economic development with an *a priori* knowledge of how events unfolded. Indicators were selected so as to describe our perception of history. This is legitimate in science - usually a hypothesis is posed and then empirical evidence is accumulated to confirm it. The next step is the search for counter examples that either refine the original hypothesis or lead to its reformulation. In either case, the benefit of hindsight offers the advantage of selecting and distilling invariants out of the universe of human and natural phenomena that remain stable over the period under investigation. This is a phenomenological approach in reconstructing history and if successful it can pave the way for the formulation of a theory. In the case of technological changes and long waves, we have accumulated empirical evidence that can be described by relatively simple rules.

The invariants indicate that at the base of human activity and interaction with the environment, selection of alternatives is a regular process. Thus, although events and the particular context of human decisions differ, the patterns of changes recur as an envelope that includes individual choice in a predictable "macro-message". Our suspicion is that the pulses of human activity are apparently not simply an artifact of capitalist economic order, but rather a more basic and elementary aspect of human nature. The last example is indicative and intriguing.

Figure 7.1 shows the index of European wheat prices from 1500 to 1869. Visual inspection of the figure indicates that four distinct trend periods occurred

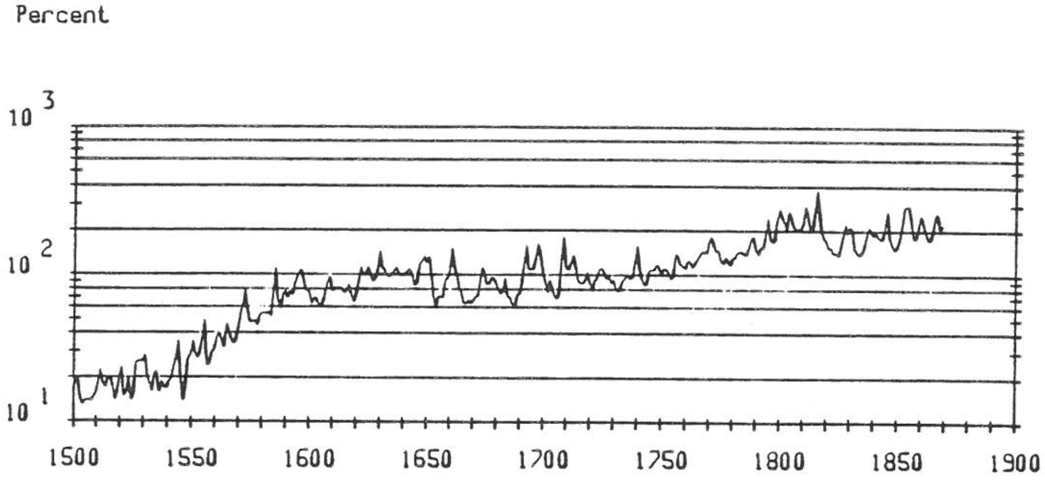


Figure 7.1 Index of European Wheat Prices (Beveridge, 1921).

between 1500 and the 1870s - a secular rise in, wheat prices during the first hundred years, followed by relatively constant prices during the seventeenth century and another period of secular rise during the eighteenth century. On average wheat prices remained constant during the first half of the nineteenth century. These secular trends are overlapped by a number of well-marked peaks and troughs of comparable relative height (with respect to the secular trend). Figure 7.2 shows the residual fluctuations of wheat prices after the elimination of the secular movements by a fifty-year moving average, together with residuals smoothed by a fifteen-year moving average that eliminates shorter cycles. The smoothed residuals are shown separately in Figure 7.3 together with the same residuals that have been smoothed twice by the fifteen-year moving average. The first price swing was initiated in 1540, peaked in 1590 and ended in 1610 with a duration of fifty years. The second swing peaked in 1635 and ended with a trough in 1675, thus it swing was somewhat longer, lasting about 65 years. The following swings coincide with the approximate fifty-year scheme given in Table 4.1 for wholesale prices in the United Kingdom and the United States.

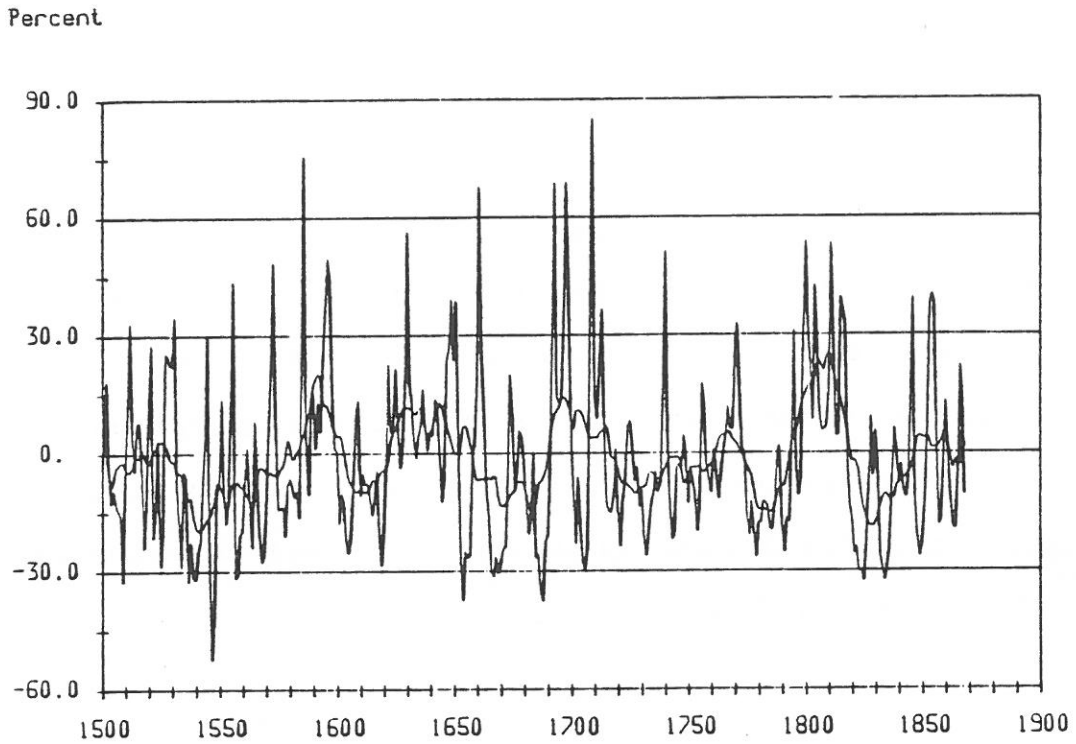


Figure 7.2 Fluctuations of European Wheat Prices.

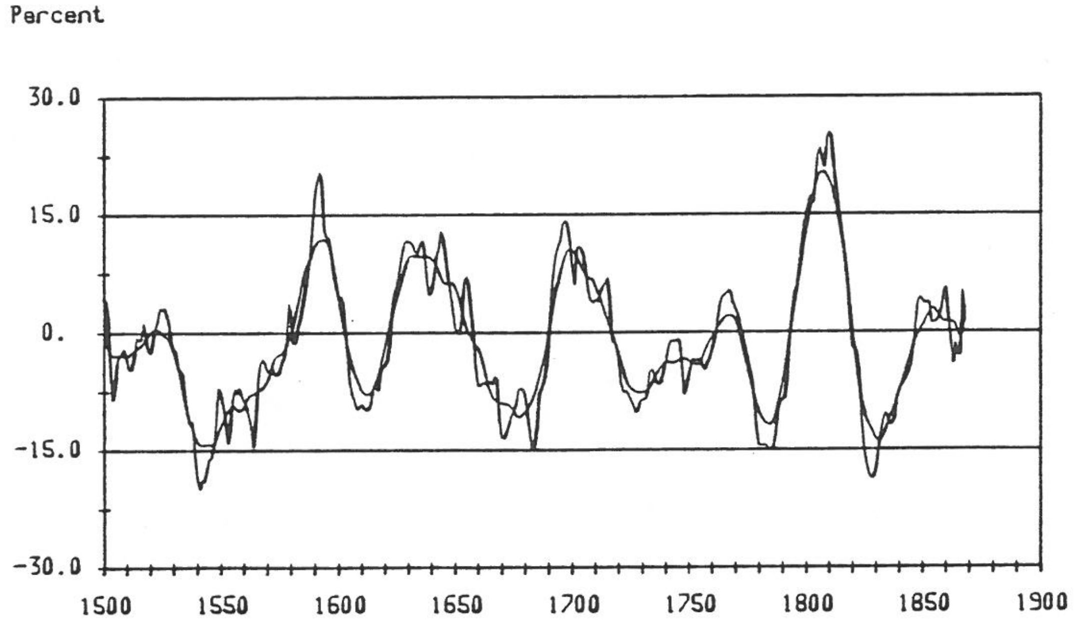


Figure 7.3 Long Wave in European Wheat Prices.

This last example illustrates the fifty-year pulses in human activities through changes in the relative prices of wheat also during the pre-industrial period. Thus, it strengthens our suspicion that the timing of the long waves and the dynamics of technology during the last two centuries was a particular realization of the more basic principles that tune the rhythm of human activities over long periods of time. It is suspicious, if perhaps a coincidence, that major wars such as the Napoleonic Wars, the Civil War and the First World War coincided with price peaks of the long swings. Even if in vain, they symbolically represent the last measure to relax the limits perceived at the end of a prosperity phase. What is attempted in warfare is achieved during the following recession through adaptation and renewal. Thereafter growth proceedes to new limits.

APPENDIX

ESTIMATION PROCEDURES

I. The Generalized Technological Substitution Model

One of the most notable models of technological substitution between two competitors was formulated by Fisher and Pry (1970). The model uses the two-parameter logistic function to describe the substitution process. The basic assumption postulated by Fisher and Pry is that once substitution of the new for the old has progressed as far as a few percent, it will proceed to completion along a logistic substitution curve

$$\frac{f(t)}{1-f(t)} = \exp(\alpha t + \beta),$$

where t is the independent variable usually representing some unit of time, α and β are constants, $f(t)$ is the fractional market share of the new competitor, and $1-f(t)$ that of the old one.

In dealing with more than two competing technologies, we have had to generalize the Fisher-Pry model since logistic substitution cannot be preserved in all phases of the substitution process. Every given technology undergoes three distinct substitution phases: growth, saturation, and decline. The growth phase is similar to the Fisher-Pry substitution, but it usually ends before full market takeover is reached. It is followed by the saturation phase, which is not logistic but which encompasses the slowing of growth and the beginning of decline. After the saturation phase of a technology, its market share declines logistically.

We assume that only one technology saturates the market at any given time, that declining technologies fade away steadily at logistic rates uninfluenced by competition from new technologies, and that new technologies enter the market and grow at logistic rates. The current saturating technology is then left with the residual market share and is forced to follow a nonlogistic path that curves from growth to decline and connects its period of logistic growth to its subsequent period of logistic decline. After the current saturating technology has reached a logistic rate of decline, the next oldest technology enters its saturation phase, and the process is repeated until all but the most recent technology are in decline. In effect, we assume that technologies that have already entered their period of market phase-out are not influenced by the introduction of new ones. The deadly competition exists between the saturating technology and all other technologies.

Let us assume that there are n competing technologies ordered chronologically in the sequence of their appearance in the market, technology 1 being the oldest and technology n the youngest, i.e., $i=1,2,\ldots,n$. This means that all technologies with indices k where $k<j$ will saturate before the technology with index j, and technologies l where $l>j$ will saturate after technology j.

Historical data of competing technologies are usually available in the form of time series for a certain period. The historical periods we have investigated range from two hundred to twenty years. Let us denote the historical time series by $g_i(t)$ where the indices $i=1,2,\ldots,n$ represent the competing technologies and t the time points of the historical period, i.e., year, month, etc. The fractional market shares of competing technologies, $f_i(t)$, are obtained by normalizing the sum of the absolute shares to one:

$$f_i(t) = \frac{g_i(t)}{\sum_j g_j(t)}.$$

By applying the linear transform of the logistic function to the fractional market shares,

$$y_i(t) = \ln\left[\frac{f_i(t)}{1-f_i(t)}\right],$$

we have n transformed time series with piece-wise linear secular trends. In fact, there are only three distinct possibilities - either a decreasing or an increasing linear trend or a phase of linear increase connected by a nonlinear saturation phase with a phase of linear decline. The oldest technology ($i=1$) always displays a declining linear trend and the youngest technology ($i=n$) an increasing linear trend (see, Figure 1.8). These linear trends can be estimated, including the increasing linear trends of technologies that enter the saturation phase during the historical period.

After the segments with linear trends of each curve $y_i(t)$ have been estimated by the method of ordinary least squares (see, Section II.) we have n equations,

$$\hat{y}_i(t) = \hat{\alpha}_i t + \hat{\beta}_i$$

where $\hat{\alpha}_i$ and $\hat{\beta}_i$ are the estimated coefficients. In order to simplify the notation we will drop the symbol "~" over the estimated statistics, so that we will denote both the observed and estimated quantities by the same symbol. However, this distinction is not necessary since from now on we will deal exclusively with estimated statistics.

These n estimated linear equations can be transformed into n logistic functions in coefficients α and β:

$$f_i(t) = \frac{1}{1+\exp(-\alpha_i t - \beta_i)},$$

where $f_i(t)$ are the estimated fractional market shares of technology i. Due to the fact that these estimated logistic functions do not capture the saturation phases and represent only growing or declining logistic trends, for some t their sum will exceed 1. Although they do not necessarily sum to 1 for all t, it is obvious that this condition must be satisfied all the time. This condition holds by definition for the observed market shares. To avoid this problem, we leave $n-1$ estimated logistic equations in their original form (i.e., as specified by coefficients α_i and β_i) and define the n-th equation as the residual, as the difference between the sum of other n-1 estimated market shares $f_i(t)$ and 1. The residual equation represents the *oldest still growing* technology and we call this technology j, such that $\alpha_j \geq 0$ where $\alpha_{j-1}<0$ and $j>1$. This chosen technology cannot be the oldest technology, i.e., $j \neq 1$, since the oldest technology is substituted by the newer technologies and, consequently, its market shares decline logistically from the start, i.e., $\alpha_1<0$.

The substitution process can be simulated over any time interval, which need not coincide with the historical period, with a limitation that it cannot precede the beginning of the historical period. Let us denote the beginning of this interval by t_b and end point by t_e.

Given the estimated coefficients α_i and β_i, and the sequence of the appearance of the technologies on the market $i=1,2,\ldots,n$, the fractional market shares are defined by

$$f_j(t) = 1-\sum_{i\neq j}\frac{1}{1+\exp(-\alpha_i t-\beta_i)}$$

and

$$f_i(t) = \frac{1}{1+\exp(-\alpha_i t-\beta_i)},$$

where $j=\{k=1,2,\ldots,n|\alpha_{k-1}<0,\alpha_k\geq 0\}$ and $i\neq j$ (i.e., $i=1,\ldots,j-1,j+1,\ldots,n$), for technology j in its transition period at time t, and where $i\neq j$, for all other technologies i. Thus, at the beginning of the time interval $[t_b,t_e]$ we have $n-1$ technologies denoted by indices $i\neq j$ which follow logistic substitution paths, at least one technology $i=1$ which leaves the market logistically, and only one technology j which forms the residual of the market, i.e. the complement of the sum of other technologies and 1. Let us denote the point in time when technology j is defined as a residual t_j. When we apply the linear transform of the logistic function to the market shares of technology j, defined above, we obtain a nonlinear function:

$$y_j(t) = \ln\left[\frac{f_j(t)}{1-f_j(t)}\right].$$

This function has negative curvature, it increases then passes through a maximum where technology j has its greatest market penetration, and finally decreases. After the slope becomes negative, the curvature diminishes for a time, indicating that $f_j(t)$ is approaching the logistic form, but then, unless technology j is shifted into its period of logistic decline, the curvature can begin to increase as newer technologies acquire larger market shares. Phenomenological evidence from a number of substitutions suggests that the end of the saturation phase should be identified with the time at which the curvature of $y_i(t)$, relative to its slope, reaches its minimum value. We take this criterion as the final constraint in our generalization of the substitution model, and from it we determine the coefficients for the j-th technology in its logistic decline.

Thus, we search for the point in time where the rate of decrease of $y_j(t)$ approximates a constant. From this point on, we set the rate of change equal to this constant and thus define a new logistic function. We approximate this point of constant slope by requiring that the relative change of slope is minimal,

$$\frac{\ddot{y}_j(t)}{\dot{y}_j(t)} = minimum$$

for $t_j \leq t < t_e$, $\ddot{y}_j(t) < 0$ and $\dot{y}_j(t) < 0$.

If this condition is satisfied, let us denote the time point when this occurs $t_{j+1} > t_j$, we determine the new coefficients for technology j

$$\alpha_j = \dot{y}_j(t_{j+1})$$

and

$$\beta_j = y_j(t_{j+1}) - \dot{y}_j(t_{j+1})t_{\ j+1}.$$

After time point t_j, technology $j+1$ enters its residual phase, and the process is repeated either until the last technology n enters the saturation phase, or the end of the time interval, t_e, is encountered.

These expressions determine the temporal relationships between the competing technologies. They have been formulated in algorithmic form. Figure A.4 shows the flowchart of the algorithm Pene that describes the logistic substitution process. A more detailed description of this procedure and the software package for the generalized logistic substitution model is given in

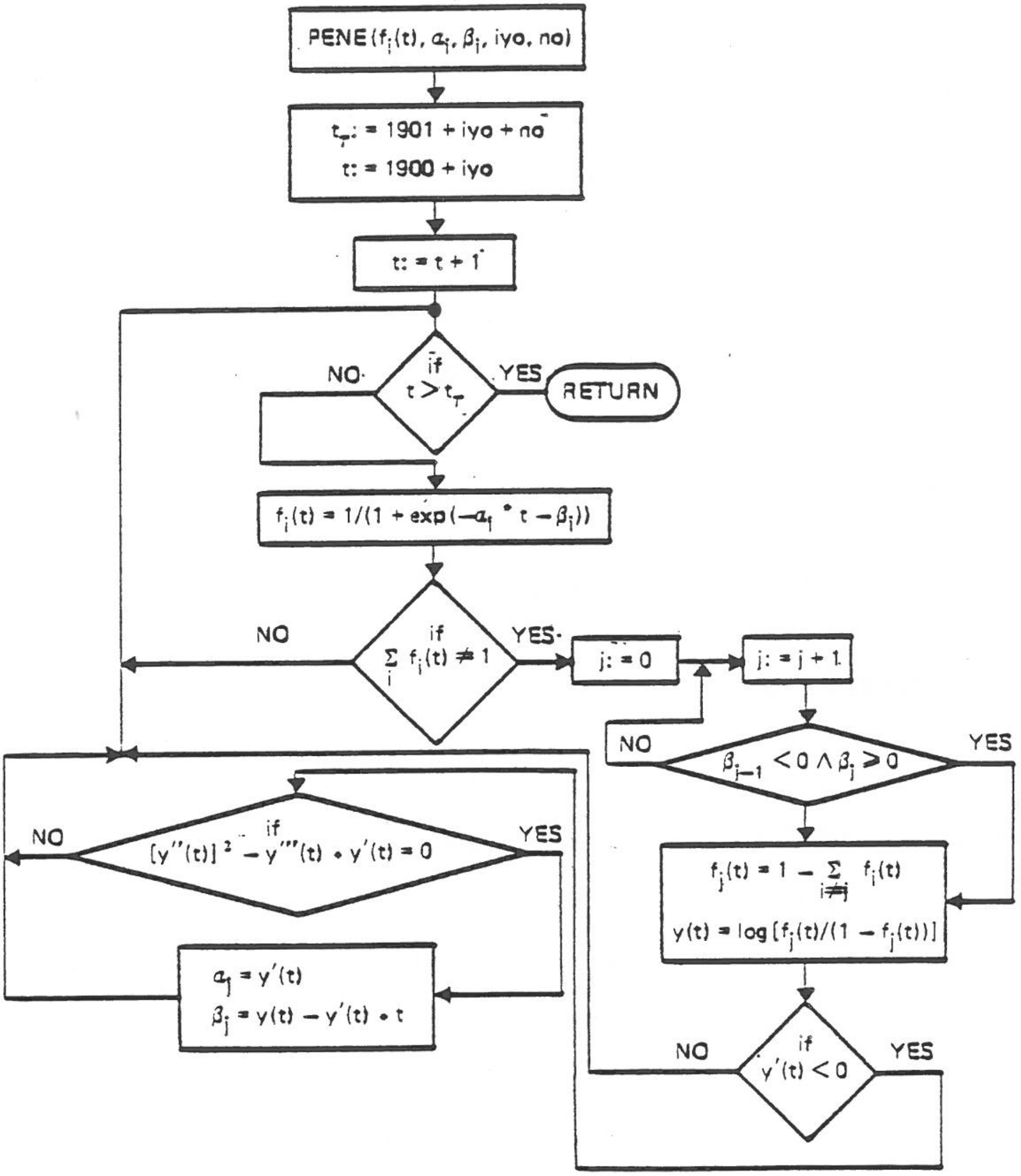


Figure A.4 Flowchart of the logistic substitution algorithm Pene.

Nakicenovic (1979 and 1980).

II. Estimation of the Logistic Growth Function

The three-parameter logistic function is defined by

$$g(t) = \frac{\kappa}{1+\exp(-\alpha t-\beta)}$$

where t is the independent variable usually representing some unit of time, κ, α and β are constants, and $g(t)$ is the value of the function at time t. The function is the generalization of the two-parameter logistic function with known parameter κ, including the special case where $\kappa=1$, and arises as the solution of the differential equation

$$\dot{g}(t) = \alpha g(t)[1-\frac{g(t)}{\kappa}],$$

where $\dot{g}(t)=dg(t)/dt$.

In empirical applications, the values of the parameters are crucial, therefore the shape of the logistic function is of greater interest than the form of its differential equation. The logistic function is monotonically increasing since the first derivative is always positive for $\alpha>0$ and it lies between two asymptotes, $g(t)=\kappa$ and $g(t)=0$. The second derivative,

$$\ddot{g}(t) = \alpha^2 g(t)[1-\frac{g(t)}{\kappa}][1-\frac{2g(t)}{\kappa}],$$

is greater than, equal to, or less than zero as $g(t)$ is less than, equal to, or greater than the value at the point of inflection. Setting $\ddot{g}(t)=0$, and solving the implicit equation for t, we obtain the point of inflection at $t=-\beta/\alpha$, and the value of the function at the inflection point $g(-\beta/\alpha)=\kappa/2$. The rate of increase of $g(t)$ is proportional to α (see the differential equation) and achieves its maximum, $\dot{g}(-\beta/\alpha)=\alpha\kappa/4$, at the inflection point and is symmetric in this point.

Thus, the three parameters represent three important attributes of the function. κ is the upper limit on the value of the function while 0 limits the function from below. α is the factor of proportionality determining the rate of increase with respect to the growth level already achieved, $g(t)$, and the amount still available for growth, $1-g(t)/\kappa$. $\exp(-\beta)$ is a location parameter, thus β acts as a location parameter; changing β while keeping other parameters fixed shifts the function horizontally but does not affect the shape of the function. For a more detailed treatment of the properties of the logistic function see Oliver (1964, 1969) and appropriate empirical applications see Brückmann (1977).

A four parameter generalization of the logistic function (see, Oliver, 1969),

$$g(t) = \gamma + \frac{\kappa}{1+\exp(-\alpha t-\beta)},$$

arises as the solution of the differential equation

$$\dot{g}(t) = \frac{\alpha}{\kappa}[g(t)-\gamma][\gamma-\kappa-g(t)].$$

The function lies between two nonzero asymptotes, $g(t)=\gamma$ and $g(t)=\gamma+\kappa$, but is otherwise unaltered since the lower asymptote simply shifts the function vertically. This generalized variant of the logistic function is useful in empirical applications because very few growth processes are initiated at infinitesimally low levels. Usually, the "start-up" of a growth process occurs at some substantial level.

The properties of the logistic function make it very attractive for describing many growth processes and consequently the estimation of its parameters has been much studied. Unfortunately, the function is not easy to fit to actual observations. Moreover, Oliver (1964) shows that "it is not the case that different methods of fitting yield much the same sort of answer provided that the data fit the function quite well. Very good fits can give quite different estimates of the parameters from one another and from those given by true least squares." The problem in empirical applications of the logistic growth processes is that the saturation level, κ, if it is not *a priori* known, is very sensitive to relatively small changes in the data base. Bruckmann (1977) notes that sometimes the addition of a single new observation can change the estimate of the saturation level (i.e., the upper asymptote κ) significantly. He suggests that in such cases it is necessary to consider other factors, such as physical constraints, in order to obtain at least an interval for acceptable values.

In those cases where the upper asymptote, κ, is known, and the lower asymptote, γ, is 0, the estimation of the other two parameters, α and β, does not present any significant problems. They can be estimated by ordinary least squares. A special case of this simpler class of estimation problems is the two-parameter function used in the logistic substitution model (see, Section I) with known upper asymptote, $\kappa=1$.

Let us denote the known value of κ by k, $k=1$ being a special case, and assume that the observations are generated by the logistic growth function with two unknown parameters α and β, and an additive, serially independent stochastic term, $u(t)$, with a normal distribution, zero mean, and constant variance σ^2, and that the values of t are predetermined, the ordinary least squares method gives the maximum likelihood estimates for the unknown parameters α and β. This assumption implies that the observations at time t are given by

$$g(t) = \frac{k}{1+\exp[-\alpha t-\beta-u(t)]} .$$

The estimation procedure is simplified by applying the linear transformation of the logistic function to observations $g(t)$,

$$\ln\left[\frac{g(t)}{k-g(t)}\right] = \alpha t + \beta + u(t),$$

to obtain transformed observations $y(t)=\ln\{g(t)/[k-g(t)]\}$ as a linear function in α and β. Now the parameters α and β are given directly by the ordinary least squares estimators $\hat{\alpha}$ and $\hat{\beta}$, and are the maximum likelihood estimates. Differentiating the logarithm of the likelihood function L, gives the following for m observations at time points $t_1,\ldots,t_m$:

$$\frac{\partial \ln L}{\partial \sigma^2} = -\frac{m}{2\sigma^2} + \frac{1}{2\sigma^4}\sum_i [y(t_i) - \alpha t_i - \beta]^2$$

$$\frac{\partial \ln L}{\partial \alpha} = -\frac{1}{\sigma^2}\sum_i \alpha [y(t_i) - \alpha t_i - \beta]$$

$$\frac{\partial \ln L}{\partial \beta} = -\frac{1}{\sigma^2}\sum_i [y(t_i) - \alpha t_i - \beta]$$

From the first equation, the maximum likelihood estimator of variance, σ^2, is evidently given by

$$\hat{\sigma}^2 = \frac{1}{m}\sum_i [y(t_i) - \alpha t_i - \beta]^2,$$

so that the maximum likelihood estimators of α and β can be derived independently of $\hat{\sigma}^2$ from the other two equations. The solution of these two equations is the same as the least squares solution and is given by:

$$\hat{\alpha} = \sum_i y(t_i)c_i$$

and

$$\hat{\beta} = \sum_i y(t_i)[\frac{1}{m} - \bar{t}c_i].$$

where $c_i = \frac{t_i \bar{t}}{\sum_i (t_i - \bar{t})^2}$ and $\bar{t} = \frac{1}{m}\sum_i t_i$ are both fixed constants.

The estimation procedure is more elaborate in the case where the value of the parameter κ is not *a priori* known. In those cases it is no longer possible to apply the linear transformation of the logistic function to observations $g(t)$ and in that way estimate the transformed observations as a linear function in α and β. We assume that the observations are generated by the logistic growth function with three unknown parameters κ, α and β, and an additive, serially independent stochastic term, $w(t)$, with a normal distribution, zero mean, and constant variance σ^2, and that the values of t are predetermined. The observations at time t are given by

$$g(t) = \frac{k}{1+\exp[-\alpha t - \beta]} + w(t).$$

Note the difference in the assumption about the stochastic term $w(t)$ compared with the earlier assumption $u(t) = \ln\left[\frac{w(t)}{\kappa - g(t)}\right]$.

To obtain the maximum likelihood estimates $\hat{\kappa}$, $\hat{\alpha}$, and $\hat{\beta}$ of the three unknown parameters κ, α, and β, we differentiate the logarithm of the likelihood function L. For m observations at time points $t_1, \ldots, t_m$ we obtain:

$$\frac{\partial \ln L}{\partial \sigma^2} = -\frac{m}{2\sigma^2} + \frac{1}{2\sigma^4}\sum_i \left[g(t_i) - \frac{\kappa}{1+\exp(-\alpha t_i - \beta)} \right]^2$$

$$\frac{\partial \ln L}{\partial \kappa} = \frac{1}{\sigma^2}\sum_i \left\{ \frac{1}{1+\exp(-\alpha t_i - \beta)} \left[g(t_i) - \frac{\kappa}{1+\exp(-\alpha t_i - \beta)} \right] \right\}$$

$$\frac{\partial \ln L}{\partial \alpha} = \frac{1}{\sigma^2}\sum_i \left\{ \frac{\kappa t_i \exp(-\alpha t_i - \beta)}{[1+\exp(-\alpha t_i - \beta)]^2} \left[g(t_i) - \frac{\kappa}{1+\exp(-\alpha t_i - \beta)} \right] \right\}$$

$$\frac{\partial \ln L}{\partial \beta} = \frac{1}{\sigma^2}\sum_i \left\{ \frac{\kappa \exp(-\alpha t_i - \beta)}{[1+\exp(-\alpha t_i - \beta)]^2} \left[g(t_i) - \frac{\kappa}{1+\exp(-\alpha t_i - \beta)} \right] \right\}$$

From the first equation, the maximum likelihood estimator of variance, σ^2, is again given by

$$\hat{\sigma}^2 = \frac{1}{m}\sum_i \left[g(t_i) - \frac{\kappa}{1+\exp(-\alpha t_i - \beta)} \right]^2 .$$

so that the maximum likelihood estimators of κ, α and β can be derived independently of $\hat{\sigma}^2$ from the other three least squares equations. Unfortunately, these three simultaneous nonlinear equations cannot be solved algebraically. The maximum likelihood estimators for the four-parameter logistic function are obtained in the same way, except that the solution of four simultaneous nonlinear equations is required. Otherwise all other aspects of the estimation procedure for the three-parameter variant are directly applicable.

In practice an iterative solution procedure is required. Oliver (1964) proposes the application of an iterative method developed by Schultz (1930), which gives the smallest sum of squared residuals, provided that it is repeated until successive adjustments become trivial and that it converges (Guest, 1961). We assume that inital values of $\hat{\kappa}$, $\hat{\alpha}$ and $\hat{\beta}$, say $\hat{\kappa}_0$, $\hat{\alpha}_0$ $\hat{\beta}_0$ are given. We will consider possible estimation procedures for obtaining such initial estimates below. The corrected values of the parameters, $\hat{\kappa}_0+\delta\hat{\kappa}$, $\hat{\alpha}_0+\delta\hat{\alpha}$ and $\hat{\beta}_0+\delta\hat{\beta}$, are derived from minimizing the approximate sum of squares

$$S = \sum_i \left[g(t_i) - g_0(t_i) - \delta\hat{\kappa}\frac{\partial g(t_i)}{\partial \hat{\kappa}_0} - \delta\hat{\alpha}\frac{\partial g(t_i)}{\partial \hat{\alpha}_0} - \delta\hat{\beta}\frac{\partial g(t_i)}{\partial \hat{\beta}_0} \right]^2$$

where $g_0(t_i)$, $\partial g(t_i)/\partial \hat{\kappa}_0$, $\partial g(t_i)/\partial \hat{\alpha}_0$ and $\partial y(t_i)/\partial \hat{\beta}_0$ are given by $g(t_i)$, $\partial g(t_i)/\partial \kappa$, $\partial g(t_i)/\partial \alpha$ and $\partial g(t_i)/\partial \beta$ with $\hat{\kappa}_0$, $\hat{\alpha}_0$ and $\hat{\beta}_0$ substituted for the unknown parameters.

This estimation procedure consists of expanding the sum of squares by Taylor's theorem, ignoring all but the first terms. Starting with the initial values, the procedure is repeated using corrected values until successive adjustments become trivial. According to Oliver (1964), this has the advantage that a unique solution is obtained independent of the initial values and that the estimates have the ordinary least squares properties, including where appropriate maximum likelihood.

Returning to the question of determining the initial values of the parameters for the iterative estimation procedure, we assume that the time series of the observed values are equally spaced, i.e., that the time points $t_1, \ldots, t_m$ are separated by time intervals or equal length. In the analysis of historical time series this is usually the case - in other applications where the observations are not equally spaced other estimates must be used. A practical possibility is to start with a given value of $\hat{\kappa}$ and than use the two-parameter logistic function to obtain corresponding values of $\hat{\alpha}$ and $\hat{\beta}$, and than iteratively change the assumed value of $\hat{\kappa}$ until sufficient accuracy is obtained for the initial estimates. An alternative procedure for obtaining the initial values of parameters for unequally spaced observations is the method of Pearl (1924) where an initial estimate of α is required to obtain $\hat{\kappa}$, $\hat{\beta}$ and the correction for $\hat{\alpha}$ iteratively. Both of these approximate methods for obtaining the initial values of the parameters have the disadvantage that the calculation effort is substantial and that the estimators have no obvious properties since they are neither least squares nor maximum likelihood estimators.

Given equally spaced observations $g(t_i)$, Bruckmann (1977), Davis (1941) and Tintner (1952) have proposed simple estimation procedures which are variants of an earlier method of Hotelling (1927). The method is based on estimating the parameters of the differential equation (see, p. 198) instead of the logistic function. The differential equation of the three-parameter logistic function can be rewritten as

$$\frac{\dot{g}(t)}{g(t)} = \alpha - \left[\frac{\alpha}{\kappa}\right] g(t).$$

Approximating $\dot{g}(t)$ by $g(t_i)-g(t_{i-1})$ when t_i are given at unit intervals, results in a difference equation

$$\frac{g(t_i)-g(t_{i-1})}{g(t_i)} = \alpha - \left[\frac{\alpha}{\kappa}\right] g(t_i).$$

Now the parameters κ and α can be estimated by linear least squares. The estimation procedure can be further simplified by using the reciprocal values of the observations instead of differences. Let us denote $1/g(t_i)$ by $z(t_i)$ and $1/g(t_{i-1})$ by $x(t_i)$. Then the method reduces to estimating the unknown parameters q and p form the transformed observations

$$z(t_i) = q + p\ x(t_i)$$

by linear least squares, where $q = \exp(-\alpha)$ and $p = [1-\exp(-\alpha)]/\kappa$. The estimates of three parameters of the logistic function are given by

$$\hat{\kappa} = \frac{1-\hat{q}}{\hat{p}},$$

$$\hat{\alpha} = \ln(\hat{q})$$

and

$$\hat{\beta} = \frac{\sum_i t_i \ln\left[\frac{g(t_i)}{\kappa - g(t_i)}\right] - \alpha \sum_i t^2_i}{\sum_i t_i}.$$

These three estimates can be used as initial values in Schultz's iterative method.

III. Estimation of the Exponential Growth Function

The exponential growth function is defined by

$$g(t) = \exp(\alpha t + \beta),$$

where t is the independent variable usually representing some unit of time, α and β are constants, and $g(t)$ is the value of the function at time t.

Under the assumption that the observations are generated by the exponential growth function with two unknown parameters α and β, and an additive, serially independent stochastic term, $u(t)$, with a normal distribution, zero mean, and constant variance σ^2, and that the values of t are predetermined, the ordinary least squares method gives the maximum likelihood estimates for the unknown parameters α and β. This assumption implies that the observations at time t are given by

$$g(t) = \exp[\alpha t + \beta + u(t)].$$

The estimation procedure is simplified by taking logarithms to obtain the linear transformation of the observations:

$$\ln[g(t)] = \alpha t + \beta + u(t).$$

Now the parameters α and β are given directly by the ordinary least squares estimators $\hat{\alpha}$ and $\hat{\beta}$ (see Section II).

IV. Moving Average

For equally spaced observations at unit intervals $t_1, \ldots, t_m$, the arithmetic moving average is defined by

$$\bar{g}(t_i) = \frac{\sum_{k=i-j}^{i+j} g(t_k)}{2j+1}$$

and the geometric moving average by

$$\bar{g}(t_i) = \exp\left[\frac{\sum_{k=i-j}^{i+j} \ln[g(t_k)]}{2j+1}\right],$$

where $i-j \geq 1$ and $i+j \leq m$.

DATA SOURCES

I. World

Primary Energy Consumption

Consumption of coal, crude oil, natural gas, hydropower and nuclear energy in the world was taken directly from Schilling and Hildebrandt (1977) for the period 1860 to 1974 and converted to GWyr/yr. For the period 1975 to 1982 consumption data were taken from BP Statistical Review of World Energy (1982) and converted to GWyr/yr after subtraction of non-energy uses of crude oil from the consumption series. Consumption of fuelwood from 1860 to 1950 was taken from Putnam (1953) and converted to GWyr/yr.

Basic Innovations

Annual number of basic innovations is from van Duijn (1983) for the period 1811 to 1971.

European Wheat Prices

Index of European wheat prices is from Beveridge (1921) for the period 1500 to 1869.

II. The United Kingdom

Primary Energy Consumption

Coal shipments from Newcastle to London are from Mitchell and Deane (1971) for the period 1655 to 1832.

For the period 1700 to 1815 coal production series is based on the coal production index compiled by Hoffmann (1940). This production index was converted into metric tons on the basis of the total production in 1854 of 65,487 thousand metric tons given in Mitchell (1975). For this period we have assumed that the apparent inland consumption of coal was equal to the production since the net exports were insignificant. Coal production and the apparent consumption (production less exports plus imports) was taken directly in metric tons from Mitchell (1975) for the period 1816 to 1922, in metric tons coal equivalent from Ministry of Power (1961) for the period 1923 to 1949, from Department of Trade and Industry (1972) for the period 1950 to 1961, from Central Statistical Office (1973) for the period 1962 to 1970, from Central Statistical Office (1983) for the period 1971 to 1977, and from Department of Energy (1983) for the period 1978 to 1982. Coal consumption was converted into GWyr/yr and included in primary energy consumption series.

The apparent oil consumption was reconstructed from oil production, imports and exports given in Mitchell (1975) and converted into GWyr/yr for the period 1856 to 1922. Consumption of crude oil, natural gas and hydropower was reconstructed from Ministry of Power (1961) and converted into GWyr/yr for the period 1923 to 1949. Hydro electricity consumption was calculated as both the primary energy equivalent (i.e., fossil fuels replacement) and as direct inputs.

Consumption of crude oil, natural gas, hydropower and nuclear energy are from Department of Energy (1983) and were converted into GWyr/yr for the period 1950 to 1982. Here again hydro electricity was calculated as primary energy equivalent and as direct electricity inputs.

Raw Steel Production

Total steel production and production by various technological processes is from Mitchell and Deane (1971) for the period 1871 to 1937, from Mitchell and Jones (1971) for the period 1938 to 1961, from Central Statistical Office (1973) for the period 1962 to 1963, from Central Statistical Office (1975) for the period 1964 to 1971 and from Central Statistical Office (1983) for the period 1972 to 1981. Data were converted into metric tons and are disaggregated by process chemistry into acid and basic and by technological production process used into Bessemer, open-hearth, basic-oxygen and electric steel.

Merchant Vessels

Number and tonnage of merchant vessels disaggregated by propulsion system into sailing, steam and motor ships is from Mitchell and Deane (1971) for the period 1788 to 1938. For the following years it was not possible to reconstruct similar series since the accounting methods were changed.

Wholesale Price Index

The index of wholesale commodity prices with base year 1930=100% is from Jastram (1977) for the period 1560 to 1938. This index was spliced with two overlapping wholesale indices from Mitchell (1975) for the period 1939 to 1968. This first of these two subindices covers the period 1914 to 1948 with base year

1929=100% and the second the period 1948 to 1969 with base year 1953=100% (the second excludes value added tax starting in 1955). Both were converted into base year 1930=100% and spliced with Jastram's index around the year 1938. It was further extended to 1981 using the wholesale price index from World Bank (1980) for the overlapping period 1950 to 1977 and wholesale price index from Central Statistical Office (1983) for the overlapping period 1974 to 1981. This extension of the index to 1981 was also converted to base year 1930=100% and spliced around the year 1968.

The quality of Jastram's index for the period prior to 1938 was also compared with well-known indices of wholesale price reproduced in Mitchell and Deane (1971) and Mitchell (1975). The Schumpeter Gilboy Price index covers the period 1661 to 1823 (Schumpeter, 1938) and is based on Beveridge's long series on prices (Beveridge et. al., 1939). The period 1790 to 1850 is covered by Gayer, Rostow and Schwrtz (1953) index and partially also by Silberling (1923), Jevons (1884) and Sauerbeck (1886). Finally, Rousseaux (1938) and Kondratieff (1925) price indices cover very long periods - 1800 to 1913 and 1780 to 1922, respectively.

Gold Price Index

Gold prices are from Jastram's gold index (Jastram, 1977) with the base year 1930=100% for the period 1560 to 1976 and from Central Statistical Office (1983) for the period 1976 to 1981.

Yield on Consols

Four overlapping sources were used to reconstruct the time series. Mitchell and Deane (1971) for the period 1756 to 1956, Mitchell and Jones (1971) for the period 1938 to 1956, Central Statistical Office (1975) for the period 1965 to 1974 and Central Statistical Office (1983) for the period 1972 to 1982.

Population

From 1701 to 1800 population estimates are based on Hoffmann's population index (Hoffmann, 1940) that includes only England, Wales and Scotland. The index was converted into population estimates in million persons on the basis of Hoffmann's index value in 1801 and the England, Wales and Scotland population of 10.686 million in 1801 (Mitchell and Deane, 1971). Population estimates for England, Wales and Scotland, Irland until 1921 and for North Irland and Republic of Irland from 1922 on are from Mitchell and Deane (1971) for the period 1801 to 1937 and from Mitchell and Jones (1971) for the period 1938 to 1960. Population of England, Wales and Scotland, and North Irland are from Central Statistical Office (1975) for the period 1961 to 1970, from Central Statistical Office (1983) for the period 1971 to 1981. Population of Republic of Irland is from World Bank (1980) for the period 1961 to 1972 and from United Nations (1981) for the period 1972 to 1981.

Gross National Product

The gross national product in current and constant prices was reconstructed from five overlapping series given in Mitchell (1975) for the period 1830 to 1961 (spliced together around the years 1855, 1870 and 1913), from Central Statistical Office (1973) for 1962 and 1963, from Central Statistical Office (1975) for the period 1964 to 1970 and from Central Statistical Office (1983) for the period 1971 to 1981. Constant gross national product was converted to 1975 prices.

Patents

Annual number of granted (sealed) patents is from Mitchell and Deane (1971) for the period 1617 to 1937, from Mitchell and Jones (1971) for the period 1938 to 1965 and from Colyer (1983, compilation based on annual publication *The Report of the Comptroller-General*, private communication) for the period 1966 to 1981.

III. The United States

Primary Energy Consumption

Consumption of Commercial Primary Energy Sources

Fuelwood consumption is from Putnam (1953) for the period 1800 to 1849 and from USHS for the period 1850 to 1970. Both time series are based on the reconstruction of fuelwood use in the United States in Reynolds and Pierson (1942) and Forest Service Report (1946). Putnam's series was adjusted to the same conversion from physical to energy units as USHS.

Consumption of bituminous, anthracite and total coal are from Putnam (1953) for the period 1800 to 1899, from Shurr and Netschert (1960) and USHS for the period 1950 to 1970, from U. S. Department of Energy (1980) for the period 1971 to 1978 and from U. S. Department of Energy (1982) for the period 1979 to 1982.

Crude oil and natural gas consumption is from Putnam (1953) for the period 1850 to 1899, from U. S. Department of Commerce (1970) for the period 1900 to 1965, from U. S. Department of Energy (1980) for the period 1966 to 1978 and from U. S. Department of Energy (1982) for the period 1979 to 1982.

Direct (mechanical) water power use is from Putnam (1953).

Hydropower is from U. S. Department of Commerce (1970) for the period 1885 to 1965, from U. S. Department of Energy (1980) for the period 1966 to 1978 and from U. S. Department of Energy (1982) for the period 1979 to 1982. Hydro electricity consumption was calculated as both fossil energy equivalent and as direct electricity inputs.

Nuclear energy consumption is from U. S. Department of Energy (1980) for the period 1960 to 1978 and from U. S. Department of Energy (1982) for the period 1979 to 1982.

All primary energy consumption series were converted from original units into GWyr/yr.

Consumption of All Primary Energy Sources

Consumption of fuelwood, draft animal feed, direct (mechanical) wind and water power, hydropower, coal, oil, natural gas and nuclear energy was taken from Fisher (1974) for the period 1850 to 1950 in five-year intervals and annually from 1950 to 1970. We have extended the time series of commercial energy sources (coal, oil, natural gas, hydropower and nuclear energy) from 1970 to 1982 from consumption series given above (Consumption of Commercial Primary Energy Sources). Hydropower is given as both fossil energy equivalent and as direct electricity inputs. Direct wind and water power is also given as both animal feed equivalent (by taking average efficiency of draft horses and mules to be four percent, i.e., by multiplying wind and water power series by factor 25) and as direct mechanical energy intputs. All consumption series were converted from original units into GWyr/yr.

Fuel Substitution in Pig Iron Smelting

Time series were taken directly from Shurr and Netschert (1960). Consumption of anthracite, bituminous coal and charcoal are given as percentages of total fuel consumed in pig iron smelting.

Raw Steel Production

Total steel production disaggregated by technological production processes (crucible, Bessemer, open-hearth, basic-oxygen and electric) is from U. S. Department of Commerce (1970) for the period 1860 to 1965, from U. S. Department of Commerce (1975) for the period 1966 to 1974 and from U. S. Department of Commerce (1981) for the period 1975 to 1980. Data were converted into metric tons from original units.

Merchant Vessels

Tonnage of merchant vessels disaggregated by propulsion system into sailing, steam and motor ships and by structural material into wood and metal is from U. S. Department of Commerce (1970) for the period 1789 to 1970. Total tonnage of the fleet is from U. S. Department of Commerce (1981) for the period 1971 to 1980.

Wholesale Price Index

The index of wholesale commodity prices with base year 1930=100% is from Jastram (1977) for the period 1800 to 1976. This index was spliced with the overlapping wholesale index given in U. S. Department of Commerce (1981) with base year 1967=100% for the period 1950 to 1981. They were spliced together around the year 1976. Jastram's index overlaps the Warren and Pearson wholesale price index (U. S. Department of Commerce, 1970) which is given for the period 1749 to 1890 and Kondratieff's index (Kondratieff, 1925) for the period 1780 to 1922.

Gold Price Index

Gold prices are from Jastram's gold index (Jastram, 1977) with the base year 1930=100% for the period 1800 to 1976. For the pericd 1977 to 1982 gold prices were compiled from U. S. Department of Interior (1980) and U. S. Department of Interior (1983) and converted into base year 1930=100%.

Population

Estimates of total resident population are from U. S. Department of Commerce (1970) for the period 1790 to 1970 and from U. S. Department of Commerce (1981) for the period 1971 to 1981.

Gross National Product

Estimates of the gross national product in current and constant prices are from Berry (1978) for the period 1789 to 1973 and from U. S. Department of Commerce (1983) for the period 1974 to 1981. Constant gross national product was converted to 1958 prices.

Patents

Annual number of granted patents is from U. S. Department of Commerce (1970) for the period 1790 to 1970, from U. S. Department of Commerce (1975) for the period 1971 to 1974, from U. S. Department of Commerce (1980) for the period 1975 to 1979 and from U. S. Department of Commerce (1981) for 1980.

BIBLIOGRAPHY

Anderson, T., 1971, The Statistical Analysis of Time Series, New York, John Wiley & Sons.

Andersson, A., 1984, Private Communication.

Berry, T., 1978, Revised Annual Estimates of American Gross National Product, Preliminary Annual Estimates of Four Major Components of Demand, 1789-1889, Bostwick Paper No. 3, University of Richmond, The Bostwick Press.

Bergson, J., 1944, Application of the Logistic Function to Bio-Assay, Journal of the American Statistical Association, 39:357-65.

Beveridge, W. et. al., 1939, Prices and Wages in England, Vol. I, London.

Beveridge, W., 1921, Weather and Harvest Cycles, December 1921, The Economic Journal, Vol. 31, 124:429-52.

Bieshaar, H. and A. Kleinknecht, 1984, Kondratieff Long Waves in Aggregate Output: An Econometric Test, CP-84-34, August 1984, Laxenburg, Austria: International Institute for Applied Systems Analysis (Forthcoming in Konjunkturpolitik, No. 5, Vol. 30, October 1984).

BP Statistical Review of World Energy, 1982, The British Petroleum Company p.l.c.

Bruckmann, G. (ed.) , 1977, Langfristige Prognosen, Möglichkeiten und Methoden der Langfristprognostik komplexer Systeme, Würzburg-Wien, Physica-Verlag.

Bruckmann, G., 1983, The Long-Wave Debate, In:
Bianchi, G., G. Bruckmann and T. Vasko (eds.), 1983, Background Material for a Meeting on Long Waves, Depression and Innovation, Siena/Florence, October 26-29, 1983, CP-83-44, August 1983, Laxenburg, Austria: International Institute for Applied Systems Analysis.

Burns, A., 1934, Production Trends in the United States Since 1870, New York: National Bureau of Economic Research, Inc.

Central Statistical Office, 1973, Annual Abstract of Statistics 1973, A publication of the Government Statistical Service, London: Her Majesty's Stationery Office.

Central Statistical Office, 1975, Annual Abstract of Statistics 1975, A publication of the Government Statistical Service, London: Her Majesty's Stationery Office.

Central Statistical Office, 1983, Annual Abstract of Statistics 1983 Edition, A publication of the Government Statistical Service, London: Her Majesty's Stationery Office.

Cipolla, C., 1976, Before the Industrial Revolution, European Society and Economy, 1000-1700, London, Methuen and Co. Ltd.

Clark, J., C. Freeman and L. Soete, 1981, Long Waves, Inventions, and Innovations, Futures, August 1981, 13:308-22.

Davis, H. T., 1941, The Analysis of Economic Time Series, Bloomington: Principia Press.

Delbeke, J., 1981, Recent Long-Wave Theories: A Critical Survey, Futures, August 1981, 13:246-57.

Department of Energy, 1983, Digest of United Kingdom Energy Statistics 1983, A publication of the Government Statistical Service, London: Her Majesty's Stationery Office.

Department of Trade and Industry, 1972, Digest of United Kingdom Energy Statistics 1972, London: Her Majesty's Stationery Office.

van Duijn, J., 1983, The Long Wave in Economic Life, London, George Allan and Unwin.

Emmens, C. W., 1941, The Dose Response Relation for Certain Principles of the Pituitary Gland, and of the Serum and Urine of Pregnancy, Journal of Endocrinology 2:194-225.

Fisher, J. C., and R. H. Pry, 1970, A Simple Substitution Model of Technological Change, Report 70-C-215, Technical Information Series, Schenectady, New York: General Electric Company, Research and Development Center.

Fisher, J. C., and R. H. Pry, 1971, A Simple Substitution Model of Technological Change, Technological Forecasting and Social Change, 3:75-88.

Fisher, J., 1974, Energy Crises in Perspective, New York, John Wiley and Sons.

Forest Service Report, 1946, A Reappraisal of the Forest Situation, Potential Requirements for Timber Products in the United States, Forest Service Report No. 2.

Forrester, J., K. Graham, P. Senge and J. Sterman, 1983, Implications for National and Regional Economic Policy, In:
Bianchi, G., G. Bruckmann and T. Vasko (eds.), 1983, Background Material for a Meeting on Long Waves, Depression and Innovation, Siena/Florence, October 26-29, 1983, CP-83-44, August 1983, Laxenburg, Austria: International Institute for Applied Systems Analysis.

Freeman, 1977, Kondratiev, Schumpeter and Kuznets: Trend Periods Revisited, In: Structural Determinants of Employment and Unemployment, 1977, Vol. II, Paris: Organization of Economic Cooperation and Development.

Garvy, G., 1943, Kondratieff's Theory of Long Cycles, The Review of Economic Statistics, November 1943, Volume 25, 4:203-20.

Gayer, A. D., W. W. Rostow and A. J. Schwartz, 1953, The Growth and Fluctuation of the British Economy 1790-1850, Vol I and II, Oxford.

van Gelderen, J. (J. Fedder), 1913, Springvloed: beschouwingen over industrieele ontwikkeling en prijsbeweging, De Nieuwe Tijd, April-June 1913, 18, 4:253-77, 5:369-84, 6:445-64.

Griliches, Z., 1957, Hybrid Corn: An Exploration in the Economics of Technical Change, Econometrica 25:501-22.

Guest, P. G., 1961, Numerical Methods of Curve Fitting, Cambridge: Cambridge University Press.

Hartman, R. and D. Wheeler, 1979, Schumpeterian Waves of Innovation and Infrastructure Development in Great Britain and the United States: The Kondratieff Cycle Revisited, Research in Economic History, 4:37-85.

Hicks, J., 1957, The Future of the Rate of Interest, T.M.S.S., London: Cambridge University Press.

Hoffmann, W., 1940, Wachstum und Wachstumsformen der Englischen Industriewirtschaft von 1700 bis zur Gegenwart, Jena. English translation:

Hoffmann, W., 1955, British Industry, 1700-1950, Oxford: Basil Blackwell.

Hotelling, H., 1927, Differential equations subject to error and population estimates, Journal of the American Statistical Association, 22:283.

Humphrey, W. and J. Stanislaw, 1979, Economic Growth and Energy Consumption in the U.K., 1700-1975, Energy Policy: March 1979, 7:29-42.

IAEA, 1983, Nuclear Power Reactors in the World, Reference Data Series No. 2, September 1983 Edition, Vienna, Austria: International Atomic Energy Agency.

Intriligator, M. D., 1978, Econometric Models, Techniques, and Applications, Oxford: North-Holland Publishing Company.

Jastram, R., 1977, The Golden Constant, New York: John Wiley & Sons.

Jevons, W. S., 1884, Investigations in Currency and Finance, London.

Jewkes, J., D. Sawers and R. Stillerman, 1958, The Sources of Invention, London: The MacMillan Press, Ltd.

Kleinknecht, A., 1981, Observations on the Schumpeterian Swarming of Innovations, Futures, August 1981, 13:293-307.

Kleinknecht, A., 1984, Innovation Patterns in Crisis and Prosperity: Schumpeter's Long Cycle Reconsidered, Academisch Proefschrift, Amsterdam: Vrije Universiteit te Amsterdam.

Kondratieff, N. D., 1925, The Major Economic Cycles, Voprosy Conjunktury, 1:28-79. German translation:

Kondratieff, N. D., 1926, Die langen Wellen der Konjunktur, Archiv für Sozialwissenschaft und Sozialpolitik, Tübingen, Verlag Von J. C. B. Mohr, Band 56:573-609. English translation:

Kondratieff, N. D., 1979, The Long waves in Economic Life, Review (Binghampton), Spring 1979, 519-62.

Kuznets, S., 1930, Secular Movements in Production and Prices, Fairfield (N.J.): Augustus M. Kelly.

Kuznets, S., 1961, Capital in the American Economy: Its Formation and Financing, New York: National Bureau of Economic Research, Inc.

Laxton, P., 1976, The geodetic and topographical evaluation of English County Maps, 1740-1840, Cartographic Journal, June 1976.

Mandel, E., 1964, The hayday of capitalism and its aftermath, Socialist Register.

Mandel, E., 1980, Long Waves of Capitalist Development, The Marxist Interpretation, Cambridge, Cambridge University Press.

Mansfield, E., 1961, Technical change and the Rate of Imitation, Econometrica 29(4).

Marchetti, C., and N. Nakicenovic, 1979, The Dynamics of Energy Systems and the Logistic Substitution Model, RR-79-13, Laxenburg, Austria: International Institute for Applied Systems Analysis.

Marchetti, C., 1979, Energy Systems - The Broader Context, Technological Forecasting and Social Change, 14:191-203.

Marchetti, C., 1981, Society as a Learning System: Discovery, Invention, and Innovation Cycles Revisited, RR-81-29, Laxenburg, Austria: International Institute for Applied Systems Analysis.

Mensch, G., 1975, Das technologische Patt, Innovationen überwinden die Depression, Frankfurt, Umschau Verlag. English Translation:

Mensch, G., 1979, Stalemate in Technology, Cambridge: Ballinger.

Miller, J., 1984, Steel Minimills, October 1984, Scientific American, 28-35.

Ministry of Power, 1961, Statistical Digest 1960, London: Her Majesty's Stationery Office.

Mitchell, B. R. and P. Deane, 1971, Abstract of British Historical Statistics, 17, Cambridge: Cambridge University Press.

Mitchell, B. R. and H. G. Jones, 1971, Second Abstract of British Historical Statistics, 18, Cambridge: Cambridge University Press.

Mitchell, B. R., 1975, European Historical Statistics, 1750-1970, London: The Macmillan Press Ltd.

Mitchell, W., 1927, Business Cycles, The Problem and Its Setting, New York: National Bureau of Economic Research, Inc.

Nakicenovic, N., 1979, Software Package for the Logistic Substitution Model, RR-79-13, Laxenburg, Austria: International Institute for Applied Systems Analysis.

Nakicenovic, N., 1980, Dynamik der Energiesubstitution, Diplomarbeit ausgeführt am Institut für Statistik und Informatik der Universität Wien unter der Betreuung von O. Prof. G. Bruckmann, Vienna Austria: Vienna University.

Nef, J., 1932, The Rise of the British Coal Industry, London, Frank Gass and Co. Ltd.

Nelder, J., 1961, The Fitting of a Generalisation of the Logistic Curve, Biometrics, 17:89-110.

Nelson, R., 1974, Neoclassical vs. Evolutionary Theories of Growth, Economic Journal, 84:886-905.

Nelson, R. and S. Winter, 1973, Toward an Evolutionary Theory of Economic Capabilities, American Economic Association, Paters and Proceedings, 63:440-49.

Nelson, R. and S. Winter, 1975, Factor Price Changes and Factor Substitution in an Evolutionary Model, The Bell Journal of Economics, 6(2):466-86.

Nelson, R. and S. Winter, 1977, In Search of a Useful Theory of Innovation, Research Policy, 6:36-76.

Oliver, F. R., 1964, Methods of Estimating the Logistic Growth Function, Applied Statistics, 13:57-66.

Oliver, F. R., 1966, Aspects of Maximum Likelihood Estimation of the Logistic Growth Function, Journal of American Statistical Association, 61:697-705.

Oliver, F. R., 1969, Another Generalization of the Logistic Growth Function, Econometrica, January 1969, 37:144-147.

Parvus, 1901, Die Handelskrisis und die Gewerkschaften, Verlag M. Ernst.

Parvus, 1908, Die kapitalistische Produktion und das Proletariat.

Pearl, R., 1924, Studies in Human Biology, Baltimore: Willians and Wilkins Co.

Pearl, R., 1925, The Biology of Population Growth, New York: Alfred A. Knopf, Inc.

Pry, R. H., 1973, Forecasting the Diffusion of Technology, Report 73CRD220, Technical Information Series, Schenectady, New York: General Electric Company, Corporate Research and Development.

Putnam, P. C., 1953, Energy in the Future, New York: D. Van Nostrand Co.

Reynolds, R. and A. Pierson, 1942, Fuel Wood Used in the United States, 1630-1930, USDA Forest Service, Circular No. 641.

Reynolds, T., 1984, Medeival Roots of the Industrial Revolution, Scientific American, July 1984, 108-116.

Richards, F., 1959, A Flexible Growth Function for Empirical Use, Journal of Experimental Botany, 10:290-300.

Robertson, T. B., 1923, The Chemical Basis of Growth and Senescence, Philadelphia: J. B. Lippincott Co.

Rostow, W. W., 1975, Kondratieff, Schumpeter, and Kuznets: Trend Periods Revisited, Journal of Economic History, 35:719-53.

Rostow, W. W., 1980, Why the Poor Get Richer and the Rich Slow Down, Essays in the Marshallian Long Period, London: The MacMillan Press, Ltd.

Rousseaux, P., 1938, Les Mouvements de Fond de l'Economie Anglaise, 1800-1913, Brussels.

Sauerbeck, A., 1886, Prices of Commodities and the Precious Metals, in J.S.S.

Schilling, H. D. and R. Hildebrandt, 1977, Primärenergie - Elektrische Energie, Die Entwicklung des Verbrauchs an Primärenergieträgern und an Elektrischer Energie in der Welt, in den USA and in Deutschland seit 1860 bzw. 1925, in W. Peters (ed.), Essen: Verlag Glückauf.

Schultz, H., 1930, The standard error of a forecast from a curve, Journal of the American Statistical Association, 25:139.

Schumpeter, E. B., 1938, English Prices and Public Finance, 1766-1822, Review of Economic Statistics.

Schumpeter, J. A., 1934, The Theory of Economic Development, Cambridge: Harvard University Press.

Schumpeter, J. A., 1935, The Analysis of Economic Change, Review of Economic Statistics, Volume 17, 2-10.

Schumpeter, J. A., 1939, Business Cycles, A Theoretical, Historical, and Statistical Analysis of the Capitalist Process, Volume I and II, New York: McGraw-Hill.

Schurr, S. and B. Netschert, 1960, Energy in the American Economy, 1850-1975, An Economic Study of its History and Prospects, Baltimore, Published for Resources for the Future, Inc. by The Johns Hopkins Press.

Scientific American, 1934, October 1934, Reprinted In:
50 and 100 Years Ago, Scientific American, October 1984, 12.

Silberling, N. J., 1923, British Prices and Business Cycles, 1779-1850, Review of Economic Statistics, 232-3.

Stapelton, E., 1976, The Normal Distribution as A Model of Technological Substitution, Technological Forecasting and Social Change, 8:325-34.

Stewart, H., 1981, Transitional Energy Policy, 1980-2030, New York: Pergamon Press.

Tintner, G., 1952, Econometrics, New York: John Wiley.

Tobin, J. and W. Nordhaus, 1972, Is Growth Obsolete?, In:
Gordon, R. (ed.), 1972, Economic Research: Retrospect and Prospect, Economic Growth, New York: National Bureau of Economic Research, Inc.

Thorp, W. and W. Mitchell, 1926, Business Annals, New York: National Bureau of Economic Research, Inc.

United Nations, 1980, Demographic Yearbook, Department of International Economic and Social Affairs, Statistical Office, United Nations, New York.

U. S. Department of Commerce, 1970, Historical Statistics of the United States, Colonial Times to 1970, Part 1 and 2, Bureau of the Census, Washington D. C.

U. S. Department of Commerce, 1975, Statistical Abstract of the United States 1975, 96th Annual Edition, Bureau of the Census, Washington D. C.

U. S. Department of Commerce, 1980, Statistical Abstract of the United States 1980, 101st Edition, Bureau of the Census, Washington D. C.

U. S. Department of Commerce, 1981, Statistical Abstract of the United States 1981, 102d Edition, Bureau of the Census, Washington D. C.

U. S. Department of Commerce, 1983, Statistical Abstract of the United States 1983-83, 103d Edition, Bureau of the Census, Washington D. C.

U. S. Department of Energy, 1980, State Energy Data Report, Statistical Tables and Technical Documentation, 1960 Through 1978, National Technical Information Services, PB80-191224, Washington D. C.

U. S. Department of Energy, 1982, Monthly Energy Review, August 1982, Energy Information Administration, DOE/EIA-0035(82/08), Washington D. C.

U. S. Department of Interior, 1980, Mineral Facts and Problems, 1980 Edition, Bureau of Mines, Washington D. C.

U. S. Department of Interior, 1983, Mineral Commodity Summaries 1983, An up-to-date Summary of 86 Nonfuel Mineral Commodities, Bureau of Mines, Washington D. C.

Verhust, P. F., 1838, In Corr. Math. et Phys, A. Quetelet, 10:113-21.

Verhust, P. F., 1845, In Nouveaux Memoires de l'Academie Royale des Sciences, des Lettres et des Beaux-Arts de Belgique, 18:1-38.

Verhust, P. F., 1847, In Nouveaux Memoires de l'Academie Royale des Sciences, des Lettres et des Beaux-Arts de Belgique, 20:1-32.

White, L., 1972, The Expansion of Technology 500-1500, In: The Fontana Economic History of Europe, Vol. 1, 1972, London.

Wilson, E. B., and J. Worcester, 1942, The Determination of L. D. 50 and Its Sampling Error in Bio-Assay, Proceedings of the National Academy of Sciences 20:79-85.

de Wolf, S., 1924, Prosperitäts- und Depressionsperioden, in: Otto Jansen, ed., Der lebendige Marxismus, Festgabe zum 70. Geburtstage von Karl Kautsky.

de Wolf, S., 1929, Het economische getij, J. Emmering.

World Bank, 1980, World Tables, The Second Edition 1980, from the data files of the World Bank.