Contributions to Economics

Arnulf Grübler

The Rise and Fall of Infrastructures

Dynamics of Evolution and Technological Change in Transport

Physica-Verlag Heidelberg

Arnulf Grübler

The Rise and Fall of Infrastructures

Dynamics of Evolution and Technological Change in Transport

With 97 Figures

Physica-Verlag Heidelberg

Series Editor
Werner A. Müller

Author
Dr. Arnulf Grübler
International Institute for Applied
Systems Analysis (IIASA)
A-2361 Laxenburg, Austria

ISBN 3-7908-0479-7 Physica-Verlag Heidelberg
ISBN 0-387-91374-2 Springer-Verlag New York

Printing: Weihert-Druck GmbH, Darmstadt
Bookbinding: J. Schäffer GmbH u. Co. KG., Grünstadt
7120/7130-543210

Preface

This book tries to merge two streams of analysis: diffusion research, a relatively recent interdisciplinary field, and the long established disciplines of transportation planning and the economic history of transport systems. After World War II, diffusion research emerged simultaneously from a number of disciplines, including sociology, geography, and the economics of technological change. For over a hundred years economic historians have provided many detailed, but often fragmented studies of the development and socio-economic impacts of individual transport systems.

At a conference held at the International Institute for Applied Systems Analysis (IIASA) in 1989 on *Diffusion of Technologies and Social Behaviour* it became clear that while many valuable theories and models have been developed, a unifying theoretical framework has not yet emerged in diffusion research. This is related to the fact that research has almost exclusively concentrated on a micro-level analysis of technological or social/cultural change. This book makes a contribution toward bridging the different fields of diffusion research. First, by providing an overview of the different theoretical streams within the discipline, and secondly by studying diffusion phenomena for technological change at the sectorial level up to the macro-level diffusion of pervasive transport systems. The economic history of transportation has provided many prominent examples of the fruitfulness of blending macro theory with a strong empirical basis at the sectorial level.

It was not the objective of this work to develop a new comprehensive theory of infrastructure development. Instead, the study aimed at combining the methodological apparatus developed within diffusion research with the empirical research tradition of the economic history of transportation. Its objective was to provide a synthetic description, covering all successive transport modes both individually in their historical evolution, and in their integration into a holistic view of the whole transport sector. The analysis is international and spans about 200 years, i.e., the developments in a number of countries are studied to highlight similarities and differences in development patterns. With such objectives, this work is both a risky and a necessarily limited effort, but it has been undertaken in the hope of providing some new perspectives for people working in, or interested in both fields.

The work has benefited from many institutions and individuals. First of all, I wish to acknowledge the intellectual and physical infrastructure of IIASA, which made this study possible. Cesare Marchetti and Nebojša Nakićenović provided many ideas, examples, and stimulating discussions. Eddie Löser and the staff of the IIASA library gave invaluable help in the never ending quest for historical data sources. Jesse Ausubel and Maximilian Posch read various drafts and provided detailed and useful comments. Special thanks are due to Eryl Mädel and Susan Riley for all their efforts at casting sheer endless sentences into English, and to Eva Delpos and Linda Foith for all their help in the preparation of this manuscript. None of them can be blamed for my errors and misinterpretations; I owe the completion of this work to their spontaneous support and continued friendship.

The study evolved out of a doctoral dissertation at the Technical University of Vienna. I thank the Institut für Finanzwissenschaft und Infrastrukturpolitik and in particular Wolfgang Blaas for hosting this endeavor. It is impossible to thank individually all the numerous institutions and individuals, who helped and contributed to this work, however, I would like to extend particular thanks to Didier Borderon and Group Planning, Shell International for positive feedback and support.

Contents

CHAPTER 1

Prologue and Introduction

> *Durch die Eisenbahnen wird der Raum getödtet, und es bleibt uns nur noch die Zeit übrig. Hätten wir nur Geld genug, um auch letztere anständig zu tödten!... ...Mir ist als kämen die Berge und Wälder aller Länder auf Paris angerückt. Ich rieche schon den Duft der deutschen Linden; vor meiner Thüre brandet die Nordsee.*
>
> Heinrich Heine, *Lutetia*, 1843

Human activities take place simultaneously in *space* and *time*. A *sequence* of activities at different coordinates of the *spatio-temporal* framework of human action (comprehensively described in the work of Hägerstrand, 1970) is made possible by *transportation*. Space and time thus define the human *activity range*, with transportation providing the basic means to cover and extend our "action radius". This is applicable whether we consider hunting and agricultural societies (Carlstein, 1982), or the presently emerging post-industrial societies.

The extent, range, and variety of activities and social contacts we experience, as well as the related costs needed (time and money) to realize them, are to a great extent, dependent on the availability and quality of our transport infrastructures and the technologies using them. The movement of people, goods, and information has consequently been a vital aspect of the functioning of all societies.

The space that can be covered by walking, and the time spent walking to, working on, and returning home from the fields defined the "economic" (i.e., the cultivated) zone around an agricultural settlement for existing industrial countries well into the 20th century and continues to do so for the rural population of most developing countries. Ancient Athens or Rome were essentially pedestrian cities, and it has been noted that their spatial size never exceeded the distance a person could cover within a maximum of 1.5 hours walking time: the diameter *intra muros*, of imperial Rome was approximately 20 *stadia*, i.e., four kilometers (km) (Putzger, 1965). The vast dimensions of still larger cities in antiquity (over two million inhabitants), like the Chinese imperial capital Ch'ang-an of the western Han and T'ang periods (206 BC to AD 8 and AD 618 to AD 907) were only possible because of the existence of an elaborate man-made transport infrastructure, i.e., canals (Schafer, 1971). Similarly, Peking provided a public wagon transportation system from a remarkably early date on (Herman and Ausubel, 1988). Herodot (Historiae V 52, 53) gives an impressive account of the size, elaborate engineering, and high quality of the transport infrastructure of the Persian Empire (including a 2,500 km imperial road between Susa and Sardes) and its vital strategic importance. The same can be said about the (military) roads of the Roman Empire, estimated to have extended over 75,000 km by the first century AD (Sax, 1920).

Empirical evidence suggests that the time devoted by an individual (on average) to transportation appears to be close to an anthropological constant: it ranges from around 1 to 1.5 hours per day, both in rural-agricultural and in urban-industrial societies. This evidence was most convincingly put forward by Zahavi, 1979 and 1981. In his Unified Mechanism of Travel (UMOT) transport model the range covered or distance travelled by an individual is defined by two constraints: the available individual *time budget* (1 to 1.5 hours per day) and the available *money budget* (around 15 percent of disposable family income).

Figures 1.1 and 1.2 and Table 1.1 document this travelling time constant. Figure 1.1 details the space-time diagram for the region of Nürnberg, Federal Republic of Germany (FRG). As family car ownership increases, the distance travelled increases also. This is the result of the transport system configuration in a particular place and point in time where cars appear to be the fastest mode of transportation. Increased accessibility of other family members to this mode of transport allows larger distances to be covered within the given time budget constraint. Table 1.1 shows that Nürnberg is not a special situation. An international survey of time allocation budgets (Szalai, 1972) shows how, on average, a 24 hour day is spent by the urban/suburban population in a number of cities/regions throughout the world. With one exception, the total travel time (commuting to and from work and other reasons for travelling, e.g., shopping or leisure) shown in Table 1.1 ranges between 1 and 1.5 hours per day. This result is all the more noteworthy, in view of the differences in the degree of economic development, access to and use of various

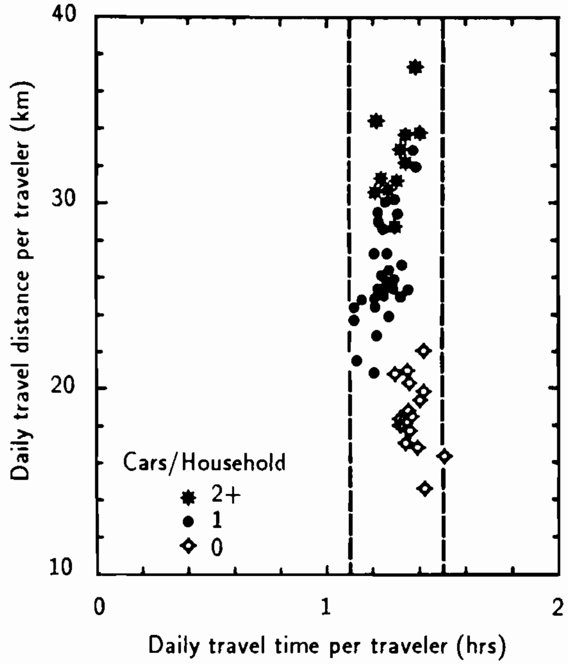


Figure 1.1. Space-time diagram of travelling in the Nürnberg region. (Source: Zahavi, 1981.)

transport means, and the different spatial settlement patterns prevailing in the countries, regions, and cities covered by the survey.

A similar invariant feature of the spatial organization of society is formulated by Lees and Hohenberg, 1988, who have observed that the distance (measured in travel time) between cities of a given size or rank tends to be more or less uniform. This occurs in addition to the regularity of the rank-size distribution of a hierarchical system of central places (see e.g., Zipf, 1949, and Carroll, 1982).

The conclusion about the static nature of the average length of total travel time that has evolved out of the cross-national and cross-sectional data discussed so far, is further corroborated by a longitudinal analysis for the USSR spanning a period of over 50 years. Figure 1.2 shows the average commuting time to and from work for men and women that has emerged from a number of time budget surveys carried out in the Soviet Union since the 1920s (summarized in Zuzanek, 1980, and updated by data from Robinson *et al.*, 1989). The static nature of commuting time ranging from 4.5 to 5.5 hours per week, persistent gender differences, as well as surprising elasticity effects (e.g., increased commuting time but a decrease in daily working

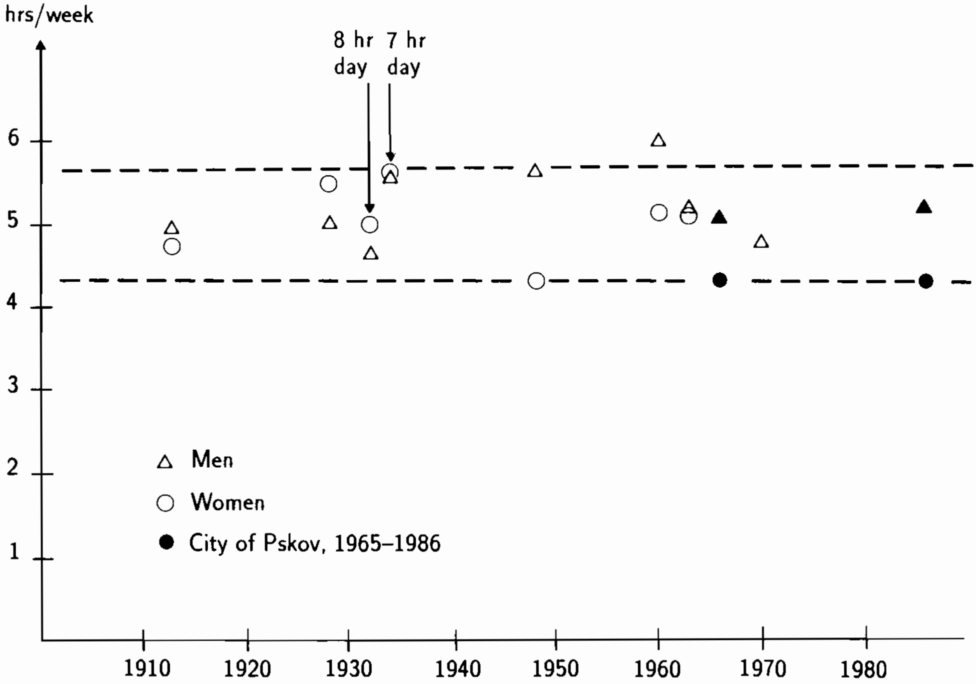


Figure 1.2. Average weekly commuting time for males and females in the USSR, hours per week.

Table 1.1. The use of time in 12 countries, hours per day. (Source: Szalai, 1972.)

	Average Time per Person Spent on Activity						
					Travel		
Country/city	Work	Household Children,..	Sleep	Leisure	Work	Non-work	Total
Belgium	4.38	5.65	8.35	4.73	0.40	0.53	0.93
Bulgaria, Kazanlik	6.05	6.04	6.97	3.55	0.68	0.80	1.48
Czechoslovakia, Olomouc	5.07	6.34	7.80	3.77	0.55	0.45	1.03
France, 6 cities	4.25	6.73	8.30	3.85	0.37	0.60	0.97
FRG, 10 districts	3.87	6.87	8.50	4.18	0.30	0.35	0.65
FRG, Osnabrück	3.63	6.60	8.34	4.68	0.27	0.70	0.97
GDR, Hoyerswerda	4.63	6.81	7.90	3.70	0.53	0.47	1.00
Hungary, Györ	5.55	6.30	7.88	3.10	0.68	0.55	1.23
Peru, Lima–Callao	3.57	5.97	8.28	4.68	0.62	0.88	1.50
Poland, Torun	4.97	5.92	7.78	4.10	0.62	0.68	1.30
USA, 44 cities	4.03	6.15	7.83	4.75	0.42	0.88	1.30
USSR, Pskow	5.65	5.43	7.70	3.77	0.55	0.92	1.47
Yugoslavia, Kragujevac	4.00	6.08	7.87	4.87	0.45	0.83	1.28

Note: Figures may not add up to 24 hours due to rounding errors.

time in the 1930s) are interesting features to note in Figure 1.2. This points to the fact that time allocation patterns and travel time budget constraints are important determinants of modal split choices, and as such, are possibly still insufficiently recognized variables in most transport analyses and models.

If we accept the notion of a time budget constraint taking the form of an invariant over time and between different cultures, then what determines the possible range that can be covered by an individual? The answer is availability, costs, performance, quality of transport infrastructures, and the technologies available to make use of these infrastructures. The most decisive performance or quality criterion for the transport system with respect to defining the realizable (spatial) range of human action is probably *speed.** Speed that exceeds the limits of human capabilities is provided by access to *infrastructure and technology*, with increasing degrees of sophistication ranging from riding horses to flying in a jet aircraft.

Marchetti, 1987, sees man as a *territorial animal* that under certain time and budget constraints will try to maximize the territory covered or "controlled" (i.e., his/her range). Transport technologies and infrastructures thus allow for the "control" of territory and the expansion of the range beyond that defined by our inherent human "performance levels". The extension of the space-time framework of human activities is made possible by the widespread application (*diffusion*) of *innovations* that take the physical form of transport and information infrastructures and the technological devices that allow us to make use of them. It is these innovations that are the subject of this study.

Particularly since the onset of the industrial revolution, many innovations have appeared that have consequently enjoyed widespread application (or failure, as illustrated for example, by the history of Zeppelins). Although generally technical in nature, innovations are all intrinsically interrelated to organizational and social adaptation processes. Once innovations appear and demonstrate potential technical and economic viability, they are put forward for societal "testing". They are either refused or accepted, and in the case of acceptance, start to *diffuse* into the economic and social environment, *interacting* with existing techniques which satisfy the same basic human need (covering range). If innovations prove to be better adapted to the technical, economic, and social requirements (boundary conditions) imposed by a society and its economy, they will start to replace existing techniques and practices.

---

* For an unconventional and critical philosophical discussion of "dromological progress" see Virilio, 1977 and 1978.

Throughout history, the availability of new transport modes, which could significantly alter the spatial and temporal activity range of people, had disruptive effects both on individuals and on the spatial organization of society. This fact is probably best illustrated by the quote from Heinrich Heine given above, which reflects the revolutionary transformation brought about by the first railway connection between Germany and France. Despite the fact that such revolutions were in the early stages often alarming to the individual, sometimes even leading to opposition or violent rejection, they nevertheless proved vital for the long-term development of industrialized economies and the resulting transformations in the spatial division of labor.

Study Outline

The dynamic picture that emerges in analyzing when and how innovations in the transport sector appear and become accepted by society, as manifest in their growth and rate of interaction (replacement or substitution) with existing techniques, is the main subject we want to describe within a *historical* and especially within a *quantitative* framework.

The objective of this study is to integrate a large body of literature in the areas of technological change and diffusion, evolutionary economics, and long-wave research, and to embed these concepts into a holistic view of the long-term developments in transportation. The analysis aims at integrating a large body of quantitative material on the development of individual infrastructures for a number of countries. The wealth of available material led us to concentrate on one specific area of technological change and infrastructures: the transport system; a few examples from other areas are also given mainly for didactical purposes. Because of data availability and international comparability, the study deals primarily with long-distance transport systems and infrastructures. However, the formal analysis framework and the conclusions derived from our analysis of long-distance transportation are, to a large extent, also applicable for the analysis of short-distance (local) transport.

It is not the aim of this study to develop a new integrative theory on long-term infrastructure development, but to reconcile existing works, analyses, and data within a consistent descriptive framework, and in particular to take into account the self-referential qualities of different systems. Thus, the study concentrates on a comparison of the *dynamics of change* and generally refrains from (normative) comparisons of transport systems in individual countries on absolute levels. As such, the study aims at assembling comprehensive evidence for regular and recurring patterns of technological change in the area of transportation, without, however, being exhaustive. A particular effort is devoted to demonstrating that in the area of technological change in transportation, we are dealing with technologically interrelated and international phenomena: the development of particular systems is characterized by interdependence and cross-enhancing at the level of technologies and by coupled dynamics and synchronization at the international level.

The material is structured into six Chapters. After the introductory Chapter, we present the methodological framework for our analysis (Chapter 2). This is provided by a number of relatively simple models originating from biology. All these models deal with the processes of growth or decline of an innovation and its *interaction* with the technological, economic, and social environment in which it is embedded. We discuss in some detail how these interaction models can be rationalized in terms of sociology, behavioral sciences, and economics. We conclude that ordered, structured evolutionary paths at the macro level are driven rather than dissipated by *diversity* in technological design, individual behavior, and *uncertainty* about the economic and social consequences of technological innovations at the micro level. In particular, the development and structural change in the transport sector will be analyzed and interpreted through the paradigm of self-organizing systems.

Chapter 3 describes quantitatively the historical development pattern of different transport infrastructures in a number of countries in chronological order, starting with canals and railways, continuing with roads, and concluding with air transport. Regularities, discontinuities, and *synchronization* in the development patterns of transport infrastructures in different countries will be outlined and examples of technological change in the devices using these infrastructures will be presented.

Chapter 4 integrates the various development trends of individual infrastructures into a holistic view of the whole transport sector. This integrative view will attempt to show some general regularities in the *structural change* of the transport system, by considering structure in terms of the *length* and *value* of the capital stock of the (physical) transport infrastructure as well as by its *performance* (output or modal split). These regularities consist of a recurring pattern in the replacement of old forms of transport by new ones in order to satisfy human needs. They appear invariant for the particular economic system for which such evolutionary processes are analyzed. This leads to a discussion of some simplified generalized propositions as regards the *driving forces* of the long-term developments in the transport sector. These are seen in the *quality* (speed and geographical coverage) of service and *performance* with respect to changing economic and social boundary conditions. An analysis of the long-term transport intensity of an economy suggests that future changes in the transport system will most likely be driven by the evolving structure of the output mix of our economies, dematerialization (a decrease in the material input per unit of output), and a shift to higher value goods and services. Finally, a brief discussion on the relationship between transport and communication is presented, and concludes that these requirements appear – at least from a long-term perspective – to be of a complementary rather than a substitutable nature. Therefore, in the presently emerging "information age", is seems unlikely that we are in a phase of discontinuity in the long-term (growth) trend in passenger transport.

After having established empirical regularities and synchronization in the historical development pattern of transport infrastructures and of technological change in their use, we suggest a model for integrating the phenomena outlined in the transport sector with the evolution of the form and direction of economic growth in general (Chapter 5). This will be done by embedding changes in infrastructures and transport technologies within the framework of discontinuities in long-term economic development (Kondratieff or long waves). The integration of infrastructure growth and technological change in the transport sector into a long-wave framework attempts to show how each major infrastructural development cycle is closely related to the difussion of a number of important technological and organizational innovations. These form a "socio-technological cluster" or "paradigm" (Perez, 1983), which dominates and drives major economic expansion periods, reaches saturation, and initiates a structural discontinuity in the long-term economic development pattern.

Because of their vital importance to society, infrastructures and transport technologies are in fact, "metaphor" (indicator) systems representative of historical economic "expansion paradigms", e.g., the railway/steam/coal or the automobile/internal combustion/oil "ages". In Chapter 6 we then draw several main conclusions from this study. First, one has to acknowledge explicitly *discontinuities* in the historical evolutionary pattern of infrastructure development. Second, we have to appreciate the substantial time constants involved in building up and decommissioning such pervasive, large-scale systems as transport infrastructures and technologies. Finally, quality of service in response to continually evolving technical, economic, and societal boundary conditions is seen as a major driving force shaping the future of the transport system. This includes, in particular, high speed, flexibility, and environmental quality, as well as institutional and organizational adjustment processes.

The study points out that of the existing long-distance transport systems, air transport corresponds best to future requirements with regards to speed, flexibility, and quality of service. Consequently, air traffic, including freight transport, is expected to grow at a rate possibly not yet fully realized by transport planners in terms of foreseeable bottlenecks in the related infrastructure. Similar performance and quality criteria will have to be satisfied by all new long-distance transport systems (e.g., high speed regional or local "metro type" links) that are introduced, complementing existing infrastructures, including, possibly as the most decisive factor, new institutional and organizational arrangements for the construction and operation of such systems.

Our transport infrastructure records the past and shapes our present and future. The current spatial configuration and the technological choices, with all their positive and negative impacts, that represent our present transport system, are the result of a long evolutionary process. Still other choices are before us, as we are apparently in a period of structural discontinuity, in which a forthcoming new socio-technical paradigm is shaped. This could lead us into a new (qualitative rather than quantitative) growth phase, possibly starting around the turn of the millennium. From a historical perspective, we have tried to develop a quantitative frame of reference to improve our understanding of the evolution of today's transport system. It is a modest and necessarily crude effort (justified only by our aim to provide a long-term and holistic picture) but we hope it will contribute to the better understanding of the complex system we call transport infrastructures, which are an expression of our history and the technical and social evolution of our society.

CHAPTER 2

Methodology

> *La société, c'est l'imitation, et l'imitation c'est une espèce de somnambulisme.*
>
> Gabriel Tarde,
> *Les lois de l'imitation*, 1895

This Chapter presents a number of mathematical models that *describe* growth and substitution processes and help to organize the empirical data base that illustrates the long-term patterns of infrastructure development and structural change in the transport sector (described in Chapters 3 and 4). All these models have a common theme: the *interaction* of a population (e.g., species) with its limiting environment, and the *interspecies* interactions (competition) among populations. By using this biological metaphor we point to the origin of these models, that is, they are all either directly taken from biology or originate from it. By analogy, these models have been applied in order to describe the dynamics of technological systems and their competition for market shares.

As a first step, these models serve as descriptive tools, and as such they do not provide insight into the driving forces and causality of the evolutionary processes they describe. They are used merely as an instrument for the systematization and quantification of long historical development patterns and to provide consistency in international comparisons. The application of such growth models to diffusion and technological substitution phenomena shows the "growth to limits" (Nakicenovic, 1984) of technological systems along regular, structured evolutionary paths. This is accompanied by a characteristic recurring pattern in their structural change, consisting of a series of replacements of old forms to satisfy human needs by new ones.

Having established the regularity of many of the processes which have shaped the present transport system, the next obvious step is to look deeper and uncover the driving forces behind this historical process. To this end, Sections 2.2 to 2.4 discuss in some detail how the temporal and spatial patterns identified are consistent, and can be rationalized based on models developed within the framework of behavioral sciences, economics, and geography. These Sections are geared to respond right from the beginning to possible objections about the "deterministic" character of the models and the processes they describe. The historical evolution described thereafter should, in fact, be understood through the paradigm of self-organizing systems. System diversity (i.e., in terms of human behavior, technological options available, and their economic and social impacts) is turbulent (basically stochastic) at the lower (for example, individual or firm) levels of aggregation, while still evolving along structured trajectories at higher levels of aggregation. Chaos thus gives birth to stability and order.

2.1. Descriptive Growth, Diffusion and Substitution Models

The main objective of this Section is to present the *formal* characteristics and properties of various growth, diffusion, and substitution models, whereas the underlying rationale for their use is treated in Sections 2.2 to 2.4.

2.1.1. Introduction into biological growth models

As mentioned in the introduction to Chapter 2, growth models are of interest for the present study as they consider growth as a process of interaction with an environment. Environment in a biological and economic sense provides *resources*: food and energy supply for biological species and the required resource inputs, and final demand for technologies. The environment, however, also imposes ultimate *constraints* to growth. In the absence of an interaction with the environment, we may simply consider evolution as an exponential growth process. However, it is precisely this *interaction* with the environment that shapes the characteristic development patterns along S-shaped trajectories, which ultimately result in a "growth to limits". It should, however, be noted that for economic and technological systems this interaction with the environment requires a more complex interpretation than traditional concepts of resource scarcity, such as that underlying for instance the Club of Rome study (Meadows *et al.*, 1972). In fact, resource scarcity proper does not appear to have directly shaped or constrained any of the growth processes in the area of transport infrastructure and technology development described in the subsequent empirical part of this study.

A quite general model of interaction in biological systems is provided by the Lotka-Volterra equation (Lotka, 1924, Volterra, 1931):

$$\frac{dN_i}{dt} = \alpha_i N_i - \frac{1}{\gamma_i} \sum_{j=1}^{n} \lambda_{ij} N_i N_j \ , \ i=1,\ldots,n \tag{2.1.1}$$

This system of differential equations describes the growth (or decline) of a species i as a function of the number (density) of the species N_i and its interaction rate with other species. The interaction in the general Lotka-Volterra equation depends on the number of species N_i and N_j as well as on the "collision rate" λ_{ij} between the species. α_i and γ_i in equation (2.1.1) are constants. In our application of this general model to the description of growth and structural change processes in the transport sector we consider three aspects of interaction:

(1) The interaction of a species N_i with its (limiting) environment in the form of a growth process.
(2) The interaction (competition) between two species N_i and N_j inside a particular ecological (or market) niche in the form of substitution models.
(3) The interaction (competition) between more than just two species in the form of a multiple (logistic) substitution model.

In cases (2) and (3) we are mainly interested in studying the interaction between "species" characterized by different levels of adaptation to the requirements imposed by the environment (the market). When a species (a technology) proves to be better fit, described by a complex vector of (technological, economic, and social) performance characteristics, to environmental (market) conditions, competition will result in a substitution pattern, with new species (technologies) replacing old ones.

Of course, the general Lotka-Volterra equation also describes other (than competitive) types of interaction between species, that in biology can also take the form of symbiotic and interdependent (predator-prey) type of relationships. In these cases the general Lotka-Volterra equation results in (mathematically) complex, oscillatory types of systems behavior, which will not be discussed here. A good discussion of the complex solutions of the Lotka-Volterra equation can be found in d'Ancona, 1939, Goell *et al.*, 1971, and Gatto, 1985. For a more philosophical discussion of the mathematics of the Lotka-Volterra equation see Peschel and Mende, 1983.

2.1.2. Simple growth models

There is a multitude of different models describing growth in interaction with a (limiting) environment. Our discussion concentrates only on those models which have also found application in the description of technological change and in economic modeling. All the functions discussed below contain (at least) three parameters, which have the following interpretation:

(i) As t tends to infinity, $y=f(t)$ approaches an upper bound that represents the level at which the growth process saturates, i.e.,

$$\lim_{t\to\infty} f(t) = K \tag{2.1.2}$$

where K is positive and finite. Furthermore we consider only* curves with $\lim\limits_{t\to-\infty} f(t)=0$.

(ii) There exists a time t_0, at which the curve has a point of inflection, i.e.,

$$\frac{d^2f}{dt^2}(t_0) = 0 \tag{2.1.3}$$

where the growth rate reaches its maximum. A growth curve is called symmetric if it is symmetric around t_0, i.e., $f(t_0-t)=f(t_0+t)$. A necessary condition for symmetry is

$$y_0 := f(t_0) = \frac{K}{2} \tag{2.1.4}$$

(iii) A third parameter, denoted by Δt, gives the length of the time interval needed to grow from 10% of K to 90% of K. More precisely, let t_p be defined by

$$f(t_p) = \frac{p}{100} K \;, \quad 0<p<100 \tag{2.1.5}$$

---

* Of course a growth process may "take off" also at initial levels >0. In such a case the original level is subtracted from the data and reintroduced thereafter in the model in the form of a constant intercept.

then Δt is given by

$$\Delta t = t_{90} - t_{10} \tag{2.1.6}$$

As a first example of a S-shaped growth curve we consider the logistic function.

(1) Three parameter logistic:

This curve is given by

$$y = f(t) = \frac{K}{1 + e^{-b(t-t_0)}} \tag{2.1.7}$$

The curve is symmetric* around t_0 and a simple calculation shows that the parameter Δt is related to the growth rate b by

$$\Delta t = \frac{1}{b}\log 81 = \frac{1}{b}4.39444915... \tag{2.1.8}$$

It was first proposed by Verhulst, 1838, as a model for human population growth and then rediscovered by Pearl, 1925, for the description of biological growth processes. The logistic function can be derived from the general Lotka-Volterra equation for the two species competition assuming a limited environment ($N_1 + N_2 = const$). The interactive mode of the growth of a species with its environment becomes also apparent by re-writing the logistic function with a linear right-hand side:

$$\log \frac{y}{K - y} = b(t - t_0) \tag{2.1.9}$$

Here the interaction between the growth achieved (available resource used, growth potential realized) y, to the growth remaining to be achieved (resources remaining to be used, remaining growth potential) $K - y$, when plotted on a logarithmic scale yields a linear function which, when plotted, highlights in particular the early and late phases of the growth process.

In addition, we may consider a three parameter logistic with data dependent weights w_k given by

---

* Other symmetric growth curves like the cumulative normal derived from probability theory (see e.g., Davies, 1979) are not discussed here because they are only gradually different from the logistic function.

$$w_k = \frac{1}{\sigma_k^2} \tag{2.1.10}$$

with

$$\sigma_k^2 = \frac{1}{K} f(t_k)[1 - f(t_k)] \tag{2.1.11}$$

Considering y_k as an observation of the random variable $Y(t_k)$, the expectation of $Y(t)$ is given by $f(t)$ and its variance by equation (2.1.11).

This choice of weights stems from a statistical interpretation of the data (see Debecker and Modis, 1986) and is dealt with in more detail in Section 2.1.5 when discussing parameter estimation algorithms and parameter uncertainties.

(2) Positively skewed S-curves:

A growth curve displaying asymmetry was proposed by Floyd, 1968, within the context of technological forecasting for the description of the evolution of technical performance or "figures of merit", with very rapid initial take-off, i.e., an asymmetrical growth pattern:

$$F(K,y) = \log \frac{y}{K-y} + \frac{y}{K-y} = bt + c \tag{2.1.12}$$

Here the inflection point, y_0, is given by

$$y_0 = f(t_0) = \frac{K}{3} \tag{2.1.13}$$

instead of $K/2$ for the logistic. The parameter Δt is given by:

$$\Delta t = \frac{1}{b}\left[\log 81 + \frac{80}{9}\right] = \frac{1}{b} 12.28333803... \tag{2.1.14}$$

(3) Gompertz function:

This non-symmetric growth function was first proposed by Gompertz, 1825, for the description of human population growth. It is given by:

$$y = f(t) = K\exp(-e^{-b(t-t_0)}) \tag{2.1.15}$$

The value at the point of inflection is given by

$$y_0 = f(t_0) = \frac{K}{e} \quad \text{where} \quad \frac{1}{e} = 0.36787944... \tag{2.1.16}$$

(e ... denotes the basis of the natural logarithm), and the parameter b is related to Δt via

$$\Delta t = \frac{1}{b}\log\frac{\log 10}{\log(10/9)} = \frac{1}{b}3.08439977... \tag{2.1.17}$$

The Gompertz function can be rewritten with a linear right hand side:

$$F(K,y) := -\log\log\frac{K}{y} = b(t-t_0) \tag{2.1.18}$$

for instance if a linear* regression algorithm is used to estimate t_0 and b.

(4) Modified exponential:

This function, which is not a genuine S-shaped curve but has been proposed several times in the diffusion literature (e.g., Coleman *et al.*, 1966), is defined by

$$y = f(t) = K(1-\mathrm{e}^{-b(t-t_0)}) \tag{2.1.19}$$

For this function the parameter t_0 does not indicate the point of inflection. The curve exhibits constantly decreasing gradients; it has no inflection point, and $f(t) < 0$ for $t < t_0$. The parameter b is related to Δt by

$$\Delta t = \frac{1}{b}\log 9 = \frac{1}{b}2.197224577... \tag{2.1.20}$$

and the "linear form" of the function is given by

$$F(K,y) = \log\frac{K}{K-y} = b(t-t_0) \tag{2.1.21}$$

---

* Because of the twofold exponentiation, non-linear least square fit algorithms do not easily converge, therefore usually linear regression of the transformed data based on equation (2.1.18) is preferred.

2.1.3. Simple substitution models

In the previous Section we used the concept of interaction as described by the general Lotka-Volterra equation in a wider sense, in that the realized growth potential is related (dependent) to the growth potential (resources) remaining to be realized (utilized). Now we turn to a more conventional interaction between biological and technological species: the case of interspecies *competition*.

In biology a competitive substitution pattern was first described by Lotka, 1924, to study the evolution of two types inside a population in which selection operates upon a species subject to Mendelian inheritance. Volterra, 1927, and 1931, as discussed in d'Ancona, 1939, analyzed the evolution of two species competing for a limited food supply, showing that if the relationship between the size of the two species differs (i.e., one species outnumbers the other – even by very small amounts) at the initial stage of the process, the system converges to a state where one species completely replaces the other. Thus, initial random fluctuations in the number of a competing species may result in a "lock-in" of the competitive pattern with one of the species replacing the other one completely. Such a concept has been proposed also for the initial selection mechanism of competing technological designs by Arthur, 1983.

As the total population in both competitive situations (the two types of species, and the two species inside the population competing for the food supply) remains constant, the evolution of competing species is renormalized in considering only the *relative* share of the competing species in the total population. This relative share evolves according to a logistic trajectory with the obvious constraint $K=1$, i.e., no species can have a higher share in the total population than 100%.

Fisher-Pry Model

Fisher and Pry, 1971, proposed a similar model to study the replacement of old by new technologies. The basic assumptions of the Fisher-Pry model are: (1) The substitution process is competitive. (2) Once substitution by the new technology has progressed as far as a few percent (i.e., "lock-in"), it will proceed until a complete takeover occurs. (3) the rate of fractional substitution F_1 is proportional to the remaining substitution $1 - F_1$ possible. Thus the model suggests a logistic substitution trajectory:

$$\frac{F_1}{1 - F_1} = \exp(a + b\,t) \tag{2.1.22}$$

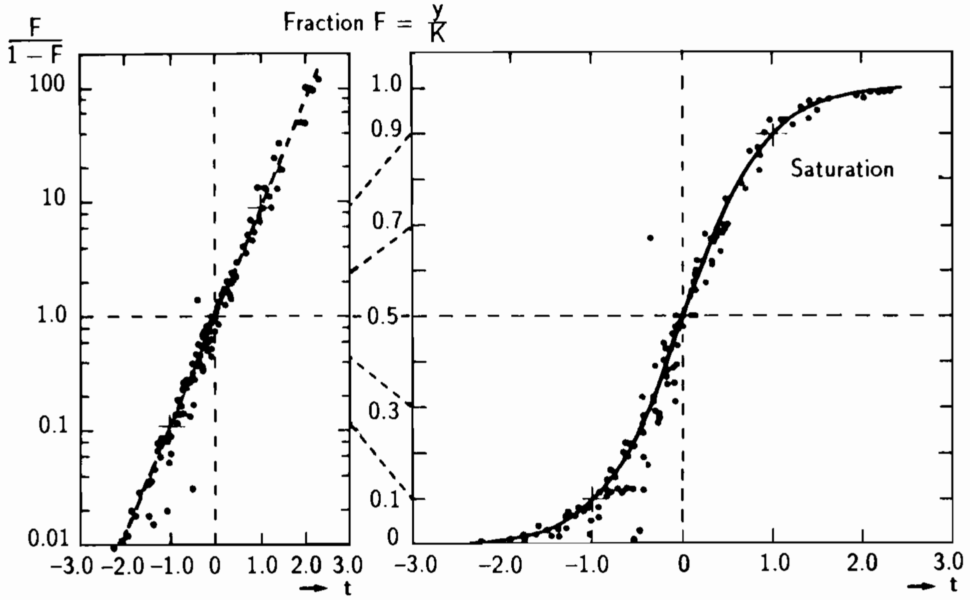


Figure 2.1.1. Life cycle in the introduction of 17 different technological innovations measuring fractional market shares. (Source: adapted from Fisher and Pry, 1971.)

where the share of the two technologies F_1 (for technology 1) and $1 - F_1$ (for technology 2) in the total market is calculated via: $F_1 = N_1 / (N_1 + N_2)$. Of course, the fractional shares may be calculated not only by measuring the number of economic or technological "species" (e.g., number of firms, number of technological objects) but also measuring their production capacity, output and so on. In equation (2.1.22) a and b are constants and t is the independent variable, usually representing time. Equation (2.1.22) is equivalent to equation (2.1.9) discussed above the only difference being that the original three parameter logistic growth curve is reduced to a two parameter (as K=1) logistic substitution curve in the Fisher-Pry model, with t_0 at F=0.5. Figure 2.1.1 shows the fractional logistic substitution curve for the introduction of 17 technological innovations (basic oxygen steel production, synthetic fibers, etc.) studied by Fisher and Pry. The curve on the left shows the linear transform $log(F/(1-F))$, i.e., market share of the new innovation divided by the market share held by the old technology on a logarithmic scale, highlighting in particular the early and late phases of the substitution process. The same model was applied by Blackman, 1971, 1972, and 1974, and numerous studies have since confirmed the descriptive power of this simple substitution model. Of course, any of the other growth functions discussed in Section 2.1.1 above can also be applied to a simple (binary) substitution case by setting $K=1$; therefore they are not discussed here again.

Sharif-Kabir Model

The only additional binary substitution model which we discuss here is the substitution model proposed by Sharif and Kabir, 1976a, which is a so-called "flexible" substitution model, in that an additional parameter is introduced to accommodate a whole range of substitution patterns from the symmetric logistic to various degrees of positively skewed (asymmetric) ones. This model, originally proposed as a substitution model (with $K=1$), becomes a growth model when the parameter K is allowed to take positive values other than one.

Using the above notation ($K=1$), this model is given by:

$$\log \frac{F_1}{1-F_1} + \gamma \frac{F_1}{1-F_1} = bt + c \ , \quad 0 \leq \gamma \leq 1 \tag{2.1.23}$$

This model contains two special cases: For $\gamma=0$ it reduces to the Fisher-Pry model while for $\gamma=1$ it corresponds to the Floyd model presented in equation (2.1.12) above (with $K=1$). For $\gamma \neq 0$ it is a non-symmetric function; the value at the inflection point, y_0, is given by

$$y_0 = f(t_0) = \frac{2}{3+\sqrt{1+8\gamma}} \ , \quad 0 \leq \gamma \leq 1 \tag{2.1.24}$$

showing that y_0 drops from 1/2 to 1/3 when γ increases from 0 to 1. Introducing the parameter t_0, equation (2.1.23) can be rewritten in the following way:

$$\log \frac{F_1}{1-F_1} + \gamma \frac{F_1}{1-F_1}$$
$$= b(t-t_0) + \log \frac{2}{1+\sqrt{1+8\gamma}} + \frac{2\gamma}{1+\sqrt{1+8\gamma}} \tag{2.1.25}$$

The parameter Δt (see Section 2.1.2) now depends on γ and is given by

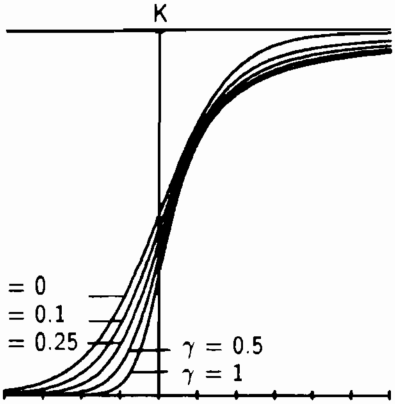


Figure 2.1.2. Sharif-Kabir substitution functions for γ=0 (logistic curve), γ=0.10, γ=0.25, γ=0.50, and γ=1 (Floyd curve). (Source: Grübler *et al.*, 1988.)

$$\Delta t = \frac{1}{b}\left[\log 81 + \frac{80}{9}\gamma\right] = \frac{1}{b}(4.39444915\ldots + \gamma 8.88888888\ldots) \quad (2.1.26)$$

Figure 2.1.2 shows some substitution curves resulting from the Sharif-Kabir model for different values of the parameter γ. This particular model offers considerable flexibility to describe a whole range of binary substitution processes, when the empirical data can not support the symmetry assumption underlying the logistic Fisher-Pry model. We have to note, however, (discussed in more detail in Section 2.2.2) that before a conclusion with respect to a deviation of a particular data set from the functional form of a substitution model can be reached, a careful analysis has to be carried out on whether the process under investigation is indeed a binary substitution process. In a majority of cases asymmetric substitution patterns are an indication that additional technologies compete on the market. In such cases the substitution process can only adequately be described by a multiple substitution model such as the Marchetti-Nakicenovic model discussed below.

2.1.4. Multiple substitution models

The main drawback of the Fisher-Pry model presented above is that it deals just with two competitors, whereas in reality more than two technologies may compete on the market. An extension of the Fisher-Pry model to a multiple substitution model was first proposed by

Marchetti and Nakicenovic, 1979*. In this model each technology undergoes three distinct phases as measured in the market share F_i – logistic growth, non-logistic saturation, and finally logistic decline. The growth and decline phases of technologies are described in the model in the same way as the Fisher-Pry model, however, two additional assumptions are made: (1) When more than two technologies compete in a market, one technology is in its (non-logistic) saturation phase defined as residual after the logistic growth/decline trajectories of the other technologies are calculated. (2) The technology that enters the saturation phase (due to the growth of newer competitors) is the oldest of the growing technologies. Thus the market share of growing/declining technologies are:

$$y_i(t) = \log \frac{F_i}{1 - F_i} = a_i + b_i t \tag{2.1.27}$$

and the market share of the saturating technology is then given by:

$$F_j = 1 - \sum_{i \neq j} F_i \tag{2.1.28}$$

The saturation phase is represented by a parabolic transition function between the linear growth and decline phases in the $\log(F/(1-F))$ transformation of the logistic substitution curves (see Figure 2.1.3 below). The model is thus nearly complete; growth and decline as well as entry into the saturation phase (defined by the growth of the $j+1$ technology) are determined. It remains to define the point where the non-logistic transition trajectory ends and the logistic decline phase for technology F_j begins. This is done by using the properties of the non-logistic transition function. This function has negative curvature, passes through a maximum (peak of market share of the technology F_j), and then starts to diminish again. The end of the saturation phase and entrance to the logistic decline phase is then defined as the point where the curvature of $y_i(t)$ relative to its slope reaches its minimum value:

$$y_j''(t) \, / \, y_j'(t) = minimum \tag{2.1.29}$$

---

* See also Marchetti, 1975, Peterka, 1977, Marchetti *et al.*, 1978, and Peterka and Fleck, 1978. For the algorithmic and computer implementation of the model see Nakicenovic, 1979. In this context we note also a simple 3-way substitution model, proposed by Sharif and Kabir, 1976b.

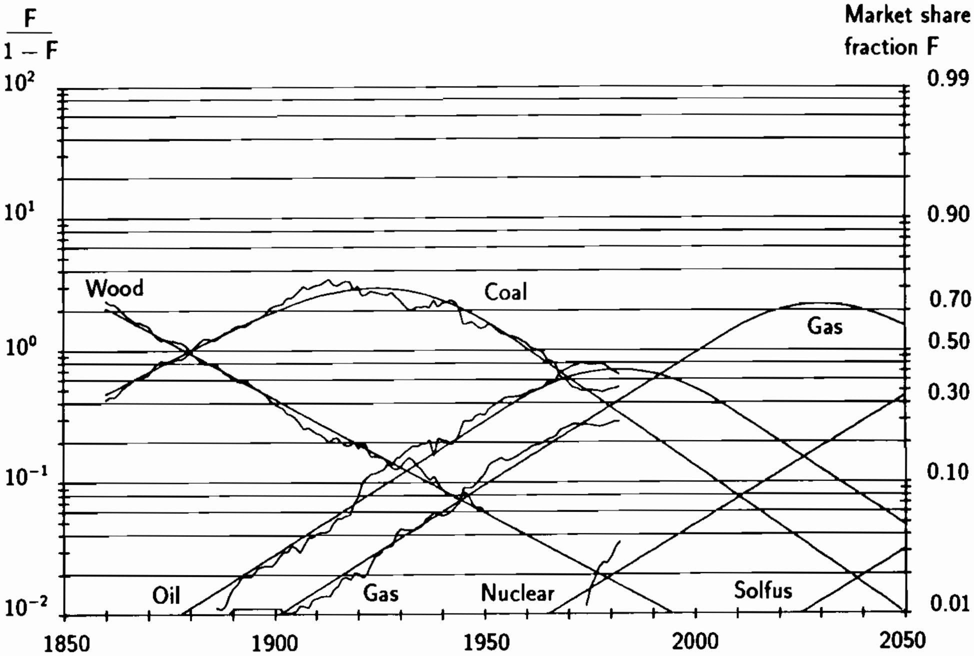


Figure 2.1.3. World primary energy substitution. (Source: Nakicenovic, 1984, after Marchetti and Nakicenovic, 1979.)

Note that y'' and y' are both negative in the region of the minimum. When this minimum condition is satisfied at time point t_{j+1}, then technology $j+1$ may in turn enter the saturation phase. The logistic decline trajectory of technology j is determined by:

$$b_j = y_j'(t_{j+1}) \tag{2.1.30}$$

$$a_j = y_j(t_{j+1}) - b_j t_{j+1} \tag{2.1.31}$$

This mechanism is continued for the successive technologies until the penultimate technology $n-1$ enters its saturation phase, leaving the market ultimately to the final winner, which in the model is assumed to be the most recently introduced technology.

Figure 2.1.3 presents such a multiple substitution case by analyzing the share of different primary energy forms in the world energy balance (Nakicenovic, 1984). The parameters of the model defining the slopes and the positioning of the various estimated substitution curves in Figure 2.1.3 are determined using an ordinary least squares regression algorithm. Because the share of nuclear energy has hardly penetrated beyond the level of a few percent of market share, its slope is assumed to be similar to the introduction of the traditional fossil energy carriers coal, oil, and natural gas. It thus represents a scenario. The technology referred to as *solfus* (to indicate that it could both be *solar* or *fusion* technology) is introduced in scenario form to analyze how the dynamics of the system would respond to the introduction of a new technology.

Before turning to the discussion on models supporting these *descriptive* models, we briefly treat the question of estimating the various parameters of the models from empirical data, as well as the statistical uncertainty related to the estimated model parameters.

2.1.5. Parameter estimation and statistical uncertainty

The main objective for using the models presented above is to map the historical evolution of transport infrastructures and technologies, both in terms of their growth and in terms of their changing *structure*. In this context it is important to discuss to what extent the various models, the parameters of which are estimated from empirical data, adequately describe the historical pattern and in particular what the statistical uncertainty of the estimated parameters (like the saturation level K) is.

The algorithms for parameter estimation are not discussed here. For the purposes of this study we have used both linear least square algorithms for parameter estimation of linear transforms of the various growth and substitution functions, as well as non-linear least squares algorithms for untransformed values. These algorithms are documented in Grübler *et al.*, 1988, and for the multiple substitution case in Nakicenovic, 1979.

With respect to parameter uncertainty it turns out that it is the parameter K, i.e., the saturation level of a particular growth process that is the most sensitive to changes in the data. For the models where this parameter is known (like in the case of market substitution models, where no technology can exceed 100% market share) parameter uncertainty is not considered a major problem. Here it suffices to rely on standard statistical measures of the goodness of fit, like R^2 or t-statistics, or leave the judgment to the human eye to assess how good a model performs in mapping the empirical data. Consequently, we are not presenting statistical measures of goodness of fit for the results obtained with these models. Sensitivity analyses with different models and estimation algorithms have been performed throughout the study and only the results with the best statistical fit are given.* From the empirical results presented in the figures in Chapters 3 and 4 one can conclude that the fit of the models is quite satisfactory, with practically all R^2 values being well above 0.9 and in a majority of cases even above 0.99.

However, when dealing with growth processes, where the asymptote is not known *a priori*, statistical uncertainty in the estimated parameters (primarily in the asymptote K) may be substantial. In principle, the minimization algorithms used provide the matrix of the second derivatives, and therefore allow the determination of the standard deviations of the values of the parameters and the corresponding confidence levels. However, this method of determining the confidence levels is not suitable, since it assumes that the errors in the parameters are normally distributed and, in addition, that they are uncorrelated. There is no compelling reason why such a restrictive assumption is warranted and as demonstrated by Debecker and Modis, 1986, there is in fact evidence that these standard statistical assumptions do not hold for estimating parameters of S-shaped growth curves. Therefore only a numerical approach, i.e., a study based on several thousands of S-curve fits on simulated data covering the different conditions for the parameters, can circumvent this problem.

Such a study, using a Monte Carlo simulation approach, was carried out by Debecker and Modis, 1986, for the three-parameter logistic function with data dependent weights w_k [see equation (2.1.10) and (2.1.11)]. This study provides tables for determining the uncertainties associated with the three parameters K, t_0, and b of a logistic growth curve based on over 35,000 simulations of randomly modified parameter values. The uncertainties, and the associated confidence levels, are given as a function of the uncertainty of the observations and the length of the historical period for which data are available. These results have been used in determining statistical uncertainties in the estimated parameters of logistic growth and diffusion processes with unknown K, and corresponding ranges for the model parameters are presented when discussing empirical results. Debecker and Modis, 1986, conclude in their study that as a rule-of-thumb the uncertainty of the parameter K (saturation level) is less than 20 percent within a 95 percent confidence level, provided at least half of the data are available** with a precision of more than 10 percent.

---

* We would like to point out that in the case of substitution, when properly analyzed (e.g., in considering the *whole* set of competing technologies via a multiple substitution model) the results of different substitution models and estimation algorithms varied only slightly and can be neglected within the context of the present study.

** That is, data are available at least up to the inflection point.

Since for other growth curves no similar information is available, uncertainty estimates for the estimated parameters cannot be presented. Consequently these models were used only in the *ex post* modeling of completed growth and diffusion processes (such as that of the railway network), where the uncertainty with respect to the asymptote of the growth process does not pose a problem and the choice of a particular growth model can be made on the basis of traditional measures of goodness of fit.

2.2. Causal Models for Temporal Diffusion and Substitution

The previous Section introduced the mathematical models used here to describe historical growth and structural change patterns in the area of transport infrastructures and transportation technologies. From a formal perspective, it may suffice to justify the use of these models if they provide satisfactory fit for the empirical data. However, it is equally desirable to identify appropriate theories that provide a rationale for the use of such models, as well as some insight into the determinants and causal factors influencing and shaping the evolutionary processes manifest in empirical data. Justification for the use of technological diffusion (growth) and substitution models is drawn from a variety of research disciplines, including sociology, social learning theory, economics, and geography, as discussed in the following Sections.

2.2.1. Innovation diffusion and adoption models

Diffusion research evolved as a discipline after World War II and studies the spread, adoption, and effects of innovations within a social system. The research tradition originates from a variety of disciplines including anthropology, (rural and medical) sociology, education research, communication theory, marketing, (agricultural) economics, and geography. The status of diffusion research has been reviewed recently at two conferences in Venice (1986) and IIASA, Laxenburg, Austria (1989).* Because of the heterogeneity of the original research disciplines it is not surprising that diffusion research has still not yet evolved into a coherent, interdisciplinary research discipline. Instead, it has been noted (Rogers, 1962) that one of the most prominent

---

* Arcangeli *et al.*, 1990 (forthcoming). Proceedings of the IIASA diffusion conference are to appear in a special issue of the journal *Technological Forecasting and Social Change*.

characteristics of diffusion research is the lack of diffusion of the research findings from one discipline to others. Consequently, there does not exist a single, universally accepted model dealing with the diffusion of innovations. However, Everett Rogers (Rogers, 1962, Rogers and Shoemaker, 1971, and Rogers, 1983) provides a number of synthesizing generalizations drawing on the findings from various research disciplines.

Following the definition given by Rogers, 1983, *diffusion* is the process by which an *innovation* is *communicated* through certain *communication channels* over *time* among the members of a *social system.*

An *innovation* may be an idea, object or practice which is perceived as new by an individual or another unit of adoption (e.g., an organization).

Communication via *communication channels* refers to how knowledge about an innovation is transferred between the members of a social system. Communication takes time, and is therefore an important aspect of innovation diffusion. In fact, all theoretical and empirical diffusion studies agree that an innovation does not instantly diffuse into a social system. Instead a typical time pattern of diffusion along a S-shaped trajectory seems to be the rule. The S-shaped pattern of diffusion appears to be a basic anthropologic phenomenon, as it is also confirmed by a large number of studies* of primitive pre-industrial societies.

The typical diffusion pattern stems from the last element of the diffusion process, i.e., from the heterogeneity of the members of a *social system* with respect to their attitudes toward an innovation. Members of a social system have a diversity of expectations about the potential benefits of an adoption decision, but are at the same time not isolated, exchanging experience (learning) and imitating the behavior of others.

It was the French sociologist Tarde, 1895, who first described the process of social change by an imitative "somnambulistic" mechanism and a S-shaped pattern in this process. It is, however, only the *aggregation* of individual behavior that portrays such a characteristic pattern. A diversity of individuals, with different expectations, value systems, communication networks, etc., *interact* with each other (sharing information and experience), and the diffusion pattern is merely an expression of this interaction between a diversity of economic agents or members of a social system.

---

* For an extensive bibliography of anthropological diffusion studies see Rogers, 1962, and Rogers and Shoemaker, 1971.

A necessary step for diffusion is therefore the transfer of knowledge and experience about an innovation from its early adopters to the rest of the population. Hägerstrand, 1952, was one of the first to recognize the importance of information flows (either via mass media or interpersonal communication) in the diffusion of innovations. Whereas mass media are more effective in communicating the *knowledge* about an innovation, interpersonal contact appears to be more effective in *influencing* the adoption decision of an individual.

Thus, information transmission is the first step in innovation diffusion. The next step concerns how that information is used in arriving at a decision. Rogers, 1983, developed a four stage model of the innovation decision process consisting of: (1) knowledge, (2) persuasion, (3) decision and implementation, and (4) (re-)confirmation.

In the knowledge stage, individual characteristics of the decision maker (social, economic, communication behavior, innovativeness) as well as his/her previous practices, perceived needs, and the general norms of the social system, act as "filter" mechanisms for incoming information. In this context Gatignon and Robertson, 1986, propose (from a marketing perspective) the following influencing parameters on information "filtering" and in consequence on the persuasion stage:

- Availability of positive information (negative information will have a greater impact than positive information).
- Credibility of information ("objective" information or information from persons with high personal or societal influence will have higher credibility).
- Consistency of information (the more consistent the higher the impact of the information).
- The type of information source (media or personal influence, with the latter being of greater influence).
- Personal characteristics (information processing style, life stage, and social integration).

After information "filtering" the persuasion stage follows and it is in this stage that the actual adoption/rejection decision is made. The obtained knowledge is evaluated in terms of the perceived characteristics of the innovation (relative advantage, compatibility, complexity, triability, and observability). These innovation characteristics are treated below in more detail when discussing the determinants of the diffusion speed. In the next stage the decision to reject or to adopt and implement an innovation is made. Finally, the decision is reassessed in that confirmation from other members of the social system is sought, which in turn may lead to the (dis)continuation of the adoption decision.

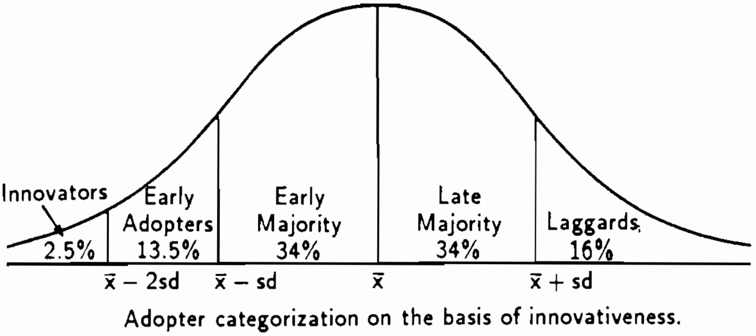


Figure 2.2.1. Distribution of adopters as a function of adoption date. (Source: Rogers, 1983.)

The above is a general model (the stages and influencing variables of which are confirmed by a large number of empirical studies) of the *individual* decision making process. However, the whole population of potential adopters is not homogeneous in terms of its innovation decision process. Consequently, Rogers, 1962, and 1983, divides the population of potential adopters according to their adoption date and categorizes them in terms of their standard deviation from the mean adoption date (Figure 2.2.1). He presents extensive empirical evidence to suggest a symmetric bell shaped curve for the distribution of adopters over time. This curve reminds us of the first derivative of the logistic growth and substitution curve presented in the previous Section and consequently the cumulative number of adopters will yield a symmetric S-shaped pattern as described by a logistic curve.

The symmetric diffusion pattern as postulated by Rogers shows *how* the members of a social system *learn* from the experience and *imitate* the behavior of the innovators (i.e., the first ones to introduce an innovation). The learning aspect is described in more detail by Casetti, 1969, and in terms of social learning by Bandura, 1977. The main trait of the argument is that the potential adopters show different degrees of resistance against an innovation as a function of their diversity in expectations, experience, etc. Resistance may be gradually overcome by appropriate stimulus, like the frequent repetition of the message (Casetti, 1969). The learning behavior of individuals is the deeper underlying cause of the symmetry of the diffusion process according to Rogers, 1983.

Decision processes for adopting an innovation can be understood as learning processes. At the beginning, when an individual is confronted with a new (learning) situation he/she makes many mistakes.

These mistakes are gradually reduced (by learning) over time because more and more information is received and acts as a stimulus. The gain in learning per trial is proportional to the product of (1) the amount already learned and (2) the amount remaining to be learned before the limit of learning is reached. Recall here that this is exactly equivalent to the $F/(1-F)$ transformation of the logistic curve discussed in the previous Section. It should be emphasized that these properties of the learning process were found in real learning situations and are confirmed by a large number of laboratory and field experiments (for a bibliography see e.g., Rogers and Shoemaker, 1971). Thus each adoption of an innovation in a social system is *equivalent to a learning trial by an individual* (Rogers, 1983). Thus, the symmetric diffusion pattern results from the way messages about an innovation are emitted and processed by social learning.

This completes the overview of diffusion models, primarily based on sociology and behavioral sciences. The only element to be discussed in more detail is the rate of adoption of an innovation, i.e., the diffusion rate (denoted by Δt in this study).

Determinants of the Diffusion Rate

Rogers, 1983, identifies five classes of variables that determine the rate of adoption of innovations.

(1) Type of innovation decisions (optional, collective or authoritative);
(2) Communication channels (external via mass media or internal through interpersonal contact, a categorization based on the work of Hamblin *et al.*, 1973);
(3) Nature of the social system (norms, degree of interconnectness).
(4) Extent of change agents promotional effort.
(5) The perceived attributes of an innovation.

However, to date there has been very little research devoted to determining the relative contribution of each of the five types of variables. For the diffusion of an innovation *inside* a larger social and economic system (like a country) it is (with the possible exception of the role of change agents) primarily the attributes of an innovation that should determine the diffusion rate.

The attributes of an innovation that influence the rate and extent to which it becomes accepted inside a social system are: *relative advantage*, *compatibility*, *complexity*, *triability*, and *observability*.

Relative advantage is the degree to which an innovation is perceived as being beneficial and, in particular, better than the idea/product/organizational form it supercedes. Perceived relative advantage may in reality be a complex vector of costs, performance, personal utility, social prestige and so on. *Compatibility* is the degree to which an innovation is considered as being consistent with existing values, past experience and individual needs and expectations. *Complexity* relates to the extent to which an innovation is perceived as being difficult to understand and to use (the learning requirement needed for its adoption). *Triability* and *observability* is the degree an innovation may be experimented with before making a final commitment on its adoption, and the degree to which the results of adoption and/or the innovation are visible to other members of the social system (in terms of learning from the experience of others and social prestige stemming from adoption). All variables, except complexity, are positively correlated to the adoption rate.

We thus conclude that the formal models introduced in Section 2.1 find an underlying causal rationale in sociological and behavioral diffusion models. In particular we recall the fact that from the diffusion perspective, the spread of a new idea, practice or product is interpreted as a *social learning* process, with consequences for the diffusion of an innovation, in particular its symmetric S-shaped pattern.

In the next Section we discuss diffusion models in economics in more detail. These models try to explain the pattern and rate of diffusion of new technologies or consumer products primarily based on the (perceived) relative advantage of an innovation, as reflected in conventional economic variables.

2.2.2. Diffusion models in economics

This Section starts with a presentation of the oldest and most influential diffusion model in economics. In a seminal contribution, Edwin Mansfield, 1961,* proposed a model explaining the adoption rates of industrial innovations, and provided empirical tests of his model for a number of innovations in various sectors.

---

* The first of this type of model was actually proposed by Griliches, 1957, who analyzed the diffusion of the use of hybrid corn seeds in various states in the USA. The approach followed is very similar to that of Mansfield both in terms of using a logistic curve to describe the diffusion patterns as well as in linking the (empirically) estimated diffusion rate to the profitability of adopting the innovation. Because this particular study dealt only with the diffusion of a single innovation and lacks a strong technological component, we prefer to discuss the Mansfield model in more detail within the present context.

We continue with an overview of the literature on first generation diffusion models and their empirical findings, and subsequent extensions of the basic model to include additional explanatory variables. As a result, a host of second generation models have been proposed, which one might group together as "equilibrium" type diffusion models. Here, diffusion is seen as a sequence of equilibria determined by changes in the economic attributes of an innovation as well as in the environment. These models are mainly discussed in terms of their relevance to the most recent, and most promising, research direction: the evolutionary type of diffusion models.

Evolutionary models, which describe diffusion as an evolutionary process under conditions of uncertainty, diversity of economic agents, and disequilibrium dynamics, try to model the complex feedback mechanisms at work at the micro level between economic agents, while still being consistent with the overall macro level ordered diffusion pattern confirmed by numerous empirical studies. Thus, these models integrate the concept of self-organizing systems as, for instance, formulated by Prigogine, 1976, into economic theory. This is illustrated by one such model in Section 2.2.3.

The Mansfield Model

Letting m_{ij} be the number of firms in the ith industry having introduced the jth innovation at time t, and n_{ij} the total number of firms, the model proposes that the number of firms at time t that will introduce the innovation at time $t+1$, λ_{ij}, is a function of: (1) the proportion of firms that have already introduced the innovation, (2) the profitability of the innovation relative to other possible investments, π_{ij}, (3) the size of the investment S_{ij}, and (4) other unspecified variables:

$$\lambda_{ij}(t) = f_i(\, m_{ij}(t) \,/\, n_{ij} \,, \pi_{ij},\, S_{ij} \,, \ldots) \tag{2.2.1}$$

The postulated relationship between the number of adopters and the number of firms having already adopted an innovation is a typical learning model where the rate of learning depends on the already accumulated knowledge. The asymptote of this learning process is reached when all potential adopters have adopted the innovation. Note the structural similarity to the linear transform of the Fisher-Pry model discussed above, where the ratio of adopters to nonadopters $F/(1-F)$, is a function of time. Consequently, Mansfield derives from equation (2.2.1) a logistic function to describe the diffusion process. Assuming that $\lambda_{ij}(t)$ can be approximated by a Taylor's expansion, which drops third and higher order terms, and assuming that the coefficient of $(m_{ij}(t)/n_{ij})^2$ in this expansion is zero and that for $t \rightarrow -\infty$, the number of adopting firms tends to zero, i.e.,

$$\lim_{t\rightarrow -\infty} m_{ij}(t) = 0 \tag{2.2.2}$$

it follows

$$m_{ij}(t) = n_{ij}\left[1 + e^{-(l_{ij}+\Phi_{ij}t)}\right]^{-1} \tag{2.2.3}$$

which is the logistic diffusion function discussed above. Note that we have retained Mansfield's original notation in this discussion.

The rate Φ_{ij} which is equivalent to the growth/substitution rate b or b_i (i.e., Δt) in Section 2.1, describes the "steepness" of the diffusion curve or is a measure of the rate of adoption (or rate of imitation in Mansfields terminology). Recall that l_{ij} is equivalent to variable a and a_i (t_0) in Section 2.1 and represents a shift parameter determining the positioning of the diffusion curve in time.

The diffusion (imitation) rate in turn depends on the profitability and the size of investments π_{ij} and S_{ij} in particular*

$$\Phi_{ij} = a_i + b\,\pi_{ij} + c\,S_{ij} + z_{ij} \tag{2.2.4}$$

where z_{ij} is an error term. The model assumes that only the intercept a_i but not the coefficients b and c vary among industries. Thus, the diffusion rate is determined by two categories of driving factors: those related to the *industry* (i.e., a_i represents an innovation index measuring the propensity towards innovation in various industries) and those related to the *innovation* (i.e., π_{ij} and S_{ij}). Empirical measures for the first category are, for instance, provided by Blackman *et al.*, 1973, for 15 industrial sectors of the USA.

In an empirical test of 12 important innovations in four industries (brewing, coal, steel, and railroad) Mansfield concludes that the logistic curve describes the diffusion pattern of innovations among firms quite well and that the (relative) profitability, in relation to other possible investments, is positively correlated (i.e., the higher the relative profitability of an investment, the faster the rate of adoption), and the required investment is negatively (i.e., the higher the investment for a particular innovation, the slower the rate of adoption) correlated with

---

* π_{ij} is defined as the average pay-out period required by the firms divided by the average pay-out period for the innovation. For relatively long lived investments, the reciprocal of the pay-out period is an approximation of the rate of return, i.e., π_{ij} is approximately equal to the average rate of return derived (*ex post*) from the innovation divided by the average rate of return firms required (*ex ante*) to justify investments. S_{ij} is defined as the percentage of the average initial investment in the innovation as a percentage of the average total assets of firms.

the diffusion rate Φ_{ij}. Mansfield found that the influence of both variables on the diffusion rate was statistically highly significant, the significance of S_{ij}, however, depends on a single observation (the tin container) in the data sample. Also significant interindustry differences (for given π_{ij} and S_{ij}) were found.

For the data sample of 12 innovations in four industries, Mansfield tested four further hypotheses regarding additional influencing variables on the rate of diffusion: The diffusion speed is negatively correlated with the average lifetime of the equipment to be replaced; the diffusion rate is positively correlated with the industry growth rate; the diffusion rate tends to increase over time (due to better communication channels, etc.); and finally, the diffusion rate depends on the time period in the (10-year) business cycle (expansion or contraction phase) when the innovation is introduced. Statistical analysis showed, however, that none of these variables had a significant influence on the diffusion rate of an innovation.

Thus, the model supports the conclusion that the rate of diffusion of a new innovation is, to a large extent, determined by the *proportion of firms already using the new technique*, the *profitability* of introducing it (compared to alternative investments), and the *size of the investment required.* Finally, significant *interindustry* differences in the rate of diffusion were observed.

In a number of subsequent studies (Mansfield 1968a, 1968b, 1977, and 1984, and Mansfield *et al.*, 1971, and 1977) the basic model was extended to include:

- Additional industrial sectors, including the chemical industry and the introduction of numerically controlled machines.
- Consideration of product innovations.
- Different measures of the diffusion process (in addition to the number of adopting firms, growth of output produced by a new industrial process, share of numerically controlled machines in new installations).
- Analysis of intrafirm diffusion rates in addition to industry diffusion rates.
- Testing a host of hypotheses associated with additional variables influencing the rate of adoption of industrial innovations.

In his analysis of the intrafirm rate of diffusion, Mansfield, 1968a and 1968b, analyzed two aspects of how particular firms respond to a new technique. First, what are the variables explaining the *time delay*, i.e., the length of time a firm waits before introducing an innovation, and second, testing a model and explaining economic variables on the *intrafirm* rate of diffusion, i.e., the pattern and driving variables of innovation diffusion *inside* different firms. It is important to consider these two aspects separately, because in the subsequent reception of the Mansfield model they have often been put together and misinterpreted, in particular with regard to the wide acceptance of the positive correlation of firm size to diffusion speed.

The main conclusions from these extensions to the original Mansfield model are:

(1) The same kind of model, i.e., logistic type, can be applied to represent diffusion among, as well as within firms, pointing at the unity and similarity between the two processes.
(2) The influence of the profitability and size of investment variables found in the original study are confirmed.
(3) The results of the effect the firm size may have on diffusion are more differentiated: whereas the size of a firm appears to significantly influence the *time* a firm waits before introducing an innovation, it "appears that small firms, once they begin, are at least as quick to substitute new techniques for old ones as their larger rivals." Company liquidity and possibilities for learning through the experience of other firms already using the innovation appear also to significantly influence the intrafirm rate of diffusion.

Before turning to a discussion of the reception and critique of the original model and the subsequent second generation of diffusion models, it may be useful to summarize the basic conclusions from the model applications. These conclusions, as put forward by Mansfield 1968a, 1968b, and 1984, and Mansfield *et al.*, 1977, also integrate the empirical findings of other studies supporting the basic model at the phenomenological and economic interpretative level; in particular those of Blackman, 1971; Hsia, 1973; Nasbeth and Ray, 1974; Simon, 1975; and Romeo, 1977. Mansfield concludes that the innovation process is essentially a learning process, characterized by a "bandwagon" or "contagion" effect. There appears to be an economic analogue to classical learning models and psychological laws, in that the extent of learning (diffusion) is a function of accumulated knowledge. Firms learn from the experience of others, who have already adopted an innovation. Furthermore the "reaction time" (time lag for adoption) and the rate of learning (adoption) are dependent on the intensity of the "stimulus", which in economic terms is represented by the profitability of the innovation.

The determinants of the rate of diffusion may therefore be summarized as follows (Mansfield, 1968b):

> Four principal factors seem to govern how rapidly an innovation's level approaches its ultimate, or equilibrium level: (1) the extent of the economic advantage of the innovation over older methods or products, (2) the extent of the uncertainty associated with using the innovation when it first appears, (3) the extent of commitment required to try out the innovation, and (4) the rate of reduction of the initial uncertainty regarding the innovation's performance.

In Mansfield, 1984, the following additional determinants were proposed:

(1) Scientific capabilities and skill level of industry (science based industries with high levels of R&D and consequently higher capabilities to evaluate proposed innovations would have higher rates of adoption).
(2) Industry and market structure, as reflected in the concentration of an industry and the dispersion of the profitability of an innovation among firms.
(3) Experience with the innovation gained outside the industry, thus reducing uncertainty via external learning.

All these determinants still constitute the backbone of any satisfactory economic modeling of innovation diffusion. They are taken up again in the rationale underlying the self-organizing model of technology diffusion and structural change discussed in Section 2.2.3.

Second Generation Diffusion Models

Since the publication of Mansfield's study, the literature dealing with diffusion in the areas of technology, economics and in marketing, has grown to such volumes that it would be outside the framework of this study to provide a detailed overview. A good compendium of earlier papers can be found in Linstone and Sahal, 1976, and concise overviews of proposed models are presented in Hurter and Rubinstein, 1978, and Mahajan and Peterson, 1985. Here we provide an overview of second generation diffusion models. We start with those that can be considered in the stream of the original models and later on structure the discussion of subsequent models around four areas of critique of the diffusion modeling approach in economics.

A number of subsequent studies have confirmed the findings about the regularity of the diffusion process, as well as the role of profitability and size of investment on the diffusion rate; in particular: Blackman, 1971, and 1974; Hsia, 1973; Simon, 1975; and Romeo, 1977, to name a few. Others, using a different analytical framework to describe the diffusion process (alternative models to the logistic or no formal model at all) have confirmed the empirical findings on the determinants of the diffusion speed postulated by first generation diffusion models, including Metcalfe, 1970; Nasbeth and Ray, 1974; and Davies, 1979, and 1980. In the international comparative study reported in Nasbeth and Ray, 1974, technological and institutional differences, and industry characteristics are also said to influence the differences in the adoption rates between countries. An exception is the study by Martino *et al.*, 1978, which could not identify any influence of an innovation's profitability on the diffusion speed, however this study is affected by a number of methodological shortcomings* and the negative conclusions should therefore be read with caution.

Finally, a number of conflicting propositions have been discussed with respect to the relationship of firm size to innovativeness and adoption rates of new technologies. The above mentioned studies confirmed Mansfield's findings, especially if we keep in mind that size appears to primarily influence the time a firm waits before starting to adopt an innovation, but that subsequently it may not necessarily be slower than earlier innovators. However, there have also been conflicting arguments that large firms concentrating on domestic markets tend to be technological laggards, as argued by Oster, 1982, in the case of the US steel industry, or that high cost firms (i.e., small companies with lower economies of scale) are more likely to adopt a new technology as argued by Reinganum, 1983. This issue is therefore still open for debate and the most likely outcome of such a discussion is that the influence of the firm size variable can go either in a positive or negative direction depending on the structure of the industry and the market, the existence of different economies of scale and so on. This problem appears to be more adequately tackled when market and industry *diversity* in diffusion models is explicitly considered, as done in Section 2.2.3.

We have structured the following discussion on second generation diffusion models along the four areas of critique that have been advanced about most of the diffusion models elaborated within economics (see e.g., Davies, 1979, Rogers, 1983, or Sahal, 1981):

(1) The mathematical properties (especially the symmetry aspect) and the adequacy of the application of the logistic model to describe the diffusion process, and the fact that the diffusion

---

* The shortcomings relate to (1) The procedure in determining the independent variables considered in the model; (2) A large number of apparent mistakes in the data gathering process from original references; (3) A restrictive definition of the appropriate market in which the innovations compete; and finally, (4) The inappropriate use of a simple (binary) diffusion model of the Fisher-Pry type for innovations which compete simultaneously on the market (i.e., in multiple substitution cases).

model is not derived functionally from an underlying economic rationale or the driving force model explaining the rate of diffusion.

(2) The nature of the adoption process, in particular the binary nature of diffusion models (both in terms of the population of potential adopters as well as in the pool of innovations available) and the static assumption on the size of the potential adopters, i.e., the competitive "niche" for an innovation (be it individuals, firms and, by extension of the model, market volume).

(3) The process innovation bias of model applications, which has led to the foundation of a second stream of innovation diffusion models (marketing models) for consumer products.

(4) The narrow definition of the group of influencing variables, the ignorance of other factors affecting diffusion as well as the fact that the models (implicitly) assume that both the economic (social) and the technological environment in which the innovation is embedded remain unchanged over time. This aspect is particularly highlighted and incorporated into a diffusion model in an important contribution by Metcalfe, 1983.

(1) Mathematical properties of diffusion models

A criticism deals with the use of a logistic model to describe diffusion patterns when it is not derived functionally from an underlying economic theory of diffusion. It is criticized that the model has been taken from a different field of reasoning, i.e., the spread of contagious diseases as applied to the dissemination of information, as well as from (social) learning theory. Secondly, the assumptions underlying the behavioral (learning) rationale of the logistic model (resulting in its symmetry around $K/2$) are considered to be too constraining. In particular the model assumes that the probability of persuasion to adopt an innovation (or the "infectiousness" of the innovation decision), as reflected in the growth rate parameter of the logistic Δt, stays constant over time.* Put in a different way, the logistic model assumes that the interaction between adopters and non adopters does not change over time. A second corollary of the logistic is that it assumes a homogeneous population, in the sense that potential adopters

---

* Note here a certain semantical difficulty in the argument. Of course the *growth rates* (i.e., the first derivative of the logistic function) vary over time in the form of a symmetric bell shaped curve with a maximum at $K/2$, resulting in the particular shape of the diffusion process over time: slow growth at the beginning, followed by very fast growth (the bandwagon effect) and finally levelling off towards the saturation level. The argument about the constant value of the diffusion rate (i.e. Δt, b, or Φ) merely refers to a perceived "lack of flexibility" (Mahajan and Peterson, 1985) of the logistic to represent data exhibiting a certain skewness (see also Davies, 1979, and Sahal, 1981, on this point).

broadly speaking share the same value system (e.g., profit maximization), and that each potential adopter is susceptible to a new innovation.

In response to this perceived "lack of flexibility" a host of alternative models have been proposed (the mathematical properties of most of them were already discussed in Section 2.1 above). These include the use of a modified exponential curve following Bass' (1969) distinction between external and internal influence driven diffusion (see Hamblin *et al.*, 1973, on the sociological rationale for these models, first proposed by Coleman *et al.*, 1966); the Floyd curve (Floyd, 1968); the Gompertz curve (e.g., Dixon, 1980); or a cumulative normal or cumulative lognormal pattern (e.g., Davies, 1979).* Other models tried to develop more comprehensive formulations with additional parameters that can accommodate a whole set of different S-shaped diffusion patterns. Examples of these type of models include the Sharif-Kabir, 1976a, model, the NSRL (Nonsymmetric Responding Logistic) model (Easingwood *et al.*, 1981), or the model proposed by Skiadas, 1986, among others.

In view of the amount of effort devoted to the development of various asymmetric diffusion models, it is surprising that the flexibility of these models (in terms that they can describe a wide range of diffusion patterns) is achieved by paying a high theoretical price. No reasoning is given by any of the model developers as to *why* a particular diffusion pattern should follow their models or what the (economic or behavioral) explanation of the additional model parameters might be. To some extent this line of research has thrown out the baby (behavioral/economic rationale of diffusion models) with the bathwater (the supposedly too constraining conditions underlying the behavioral rationale of the logistic diffusion model).

The only remaining justification for these models appears to be that they can describe *ex post*** the diffusion patterns more accurately. Whereas we do not argue that there are diffusion/substitution processes that do not conform to logistic trends (and in the empirical part of the study we use the best fit model wherever appropriate to describe historical patterns), it is somehow ironic that in the majority of cases the logistic performed better than the "more flexible" models. For instance the fit of the logistic turns out to be superior in 9 out of

---

* Davies reasons that a cumulative normal (a symmetric S-shaped curve) describes adequately the diffusion of innovations characterized by relatively large capital outlays and complex technology, whereas the (positively) skewed cumulative lognormal curve is assumed for relatively inexpensive, uncomplicated processes (often of a supplementary nature).

** *Ex ante* use of these models does not appear reasonable in view of the absence of a theoretical rationale underlying the model application.

10 cases analyzed by Sahal, 1981, and in 6 out of 8 cases analyzed by Sharif and Kabir, 1976a.

We conclude this discussion on the use of symmetric versus asymmetric diffusion models by considering one final point. Hardly any innovation is introduced into a vacuum. In fact, in the majority of innovation diffusion/substitution cases more than two technologies (innovations) compete on the market. Therefore it is entirely inadequate to arrive at conclusions with respect to a particular innovation diffusion pattern by looking at it in isolation. An extremely illustrative case is provided by Skiadas, 1986, in his comparison of various asymmetric diffusion models to the adoption of the oxygen steel process in different countries. It can be easily shown that the levelling off of the diffusion of oxygen steel (i.e., the "retarding" of diffusion, resulting in asymmetry) is simply the result of the oxygen process being replaced in turn by a newer process, namely the electric arc process (growing itself logistically). Therefore, it is completely pointless to argue in favor of a particular model in the absence of a theoretical basis for the functional form of the diffusion process and without a complete analysis of the technological and market environment (i.e., the competitive "niche") the innovation is embedded in.

The last area of criticism in relation to logistic diffusion/substitution models is the fact that the model is not derived functionally from an underlying economic rationale. This critique was certainly valid for the first generation of models. However, in the mean time there have been a number of formulations proposed to show how a particular diffusion curve can be deduced functionally from an underlying economic rationale. Davies, 1979, deduces a cumulative normal curve i.e., a symmetric diffusion curve, which he (Davies, 1980) considers equivalent to the logistic, based on the plausible assumption of a lognormal distribution of firm size. Russell, 1980, considering marketing diffusion models, argues that income distribution is lognormally distributed and if the price of a new consumer item will fall linearly, the resulting diffusion process will yield a lognormal curve. A similar line of argument is followed by Gottinger, 1985, who takes a probabilistic perspective of adoption and derives a logistic diffusion curve from a logistic probability distribution of the utility function vector of the possible entrants. Finally Tani, 1988, derives a Gompertz curve for the diffusion of industrial robots in Japan based on a rank size distribution of firm size and a size-dependent distribution of the potential benefits of adoption. The underlying rational is that the labor reduction effect of robots is higher in large firms than in small ones, due to different economies of scale. Thus both symmetric (like logistic or lognormal) and asymmetric diffusion curves (like the Gompertz) have been analytically derived from underlying economic rationales. This justifies therefore the use of these models even in those cases where data limitations (distribution of firm size, income, etc.) do not allow the inclusion of these additional variables into the diffusion model proper.

(2) The nature of the adoption process

The binary nature of most diffusion models with respect to the adopting population (by considering just adopters and non adopters) is certainly an oversimplification of the various stages of the adoption process (awareness, knowledge, etc.) formulated for instance by Rogers, 1962, and 1983. However, in an economic context such a model simplification appears justified, as we are mainly interested in the *impact* of the adoption decision with its resulting economic consequences. Within this context another critical point has been raised, namely that diffusion models tend to be pro-innovation biased and do not offer detailed insights into the mechanisms why an innovation is not adopted at all (see Rogers, 1983, on this point). Clearly this is both an empirical and philosophical problem, as unsuccessful innovations are usually very badly documented, if at all, and to date practically no diffusion researcher has pursued a Popperian falsification approach in analyzing technological change. However, we do not consider it a major deficiency in the approach as long as one is primarily interested in studying the economic *effects* of innovations.

Finally, it has been noted that diffusion models generally assume implicitly only one adoption per adopter (i.e., excluding repeat purchases), and that the models do not consider the possibility that an adopter might give up the innovation ("curing") and readopt it again at a later stage ("reinfection"). The exclusion of repeat purchases is certainly a drawback for the marketing analysis of the diffusion of consumer durables, where most models (e.g., Bass, 1969) are concerned only with initial purchases. There appear to be only two pragmatic ways out of this situation, the first is not to consider flow variables (like sales), but stock variables (like items in use or per capita values) instead. Thus, only the increase in stock would be used to measure the diffusion of a new product into the market. The second more difficult, but at the same time more realistic, possibility would be to model initial purchase and replacement demand separately. This is more plausible, as the time constant involved in diffusion and the resulting annual sales might be very different from the annual replacement demand, which depends on the size of the age cohorts of the product and its average life time as for instance represented dynamically by a "death curve" (see e.g., Marchetti, 1983, for such a model for the Japanese car market). A diffusion model incorporating replacement, in addition to expansion investments, is discussed in Section 2.2.3 below.

The critique on the binary nature of the population of innovations treated by most diffusion models is to be taken seriously. In fact, most models cannot even be considered to be of a binary nature in dealing *strictu sensu* with the case of diffusion, e.g., in analyzing the number of firms adopting a particular technique. Such models ignore the fact that potential adopters do not introduce the innovation into a vacuum, but instead replace existing techniques, equipment, etc. Therefore the process of technological change should in most cases be considered as a substitution rather than a diffusion phenomenon. This argument was particularly convincingly put forward by Sahal, 1981, and Mahajan and Peterson, 1985, among others.

The pioneering work in overcoming this deficiency in diffusion models is certainly due to Fisher and Pry, 1971 (and also Blackman, 1971), who analyzed diffusion/substitution on the basis of measuring the relative market share F of old versus new innovations competing in a market. This model was discussed in more detail in Section 2.1 above. Note, that with this model no restriction on the ultimate market volume of the innovation is imposed, which can continuously change over time, as only the relative evolution of market shares is considered. The same holds true for the multiple competition model of Marchetti and Nakicenovic, 1979* (discussed in Section 2.1 above).

The development of (relative) market share models as discussed above resolves most of the problems in dealing with changing market sizes, however, it leaves one open question, namely what is the impact of the diffusion of an innovation on market growth? The answer is discussed below in considering Metcalfe's, 1983, model where diffusion is considered as an adjustment process between two equilibria (both in terms of technology as well as in market volume) and where the increase in market volume is determined by the shift in the demand/supply curves as a result of changing prices due to the introduction of an innovation. With respect to the critique on the constant geographical size of the market niche for diffusion assumed in most models, we mention that temporal diffusion models may also be applied to the spatial spread of innovations (see the integrated time/space diffusion model discussed in Section 2.4 below).

---

* Peterka, 1977, provides an economic model explaining the dynamics of multivariate competition, based on production costs and specific investment differentials as well as a total growth rate factor.

(3) Process innovation bias

The critique on the process innovation bias* of diffusion models has led to a second stream of innovation diffusion models (marketing models) for consumer products. The formal descriptive characteristics of these models is very similar to the S-shaped models of technological innovation diffusion and substitution, including a large variety of models developed to represent asymmetric S-shaped diffusion patterns.** However, marketing diffusion models differ somewhat in their underlying behavioral rationale and the driving forces of the diffusion process. In addition to the relative advantage and costs of a new consumer product, marketing models also emphasize other factors relating to the potential adopters perception of an innovation and its attributes such as product complexity, its compatibility with existing consumer experience, its triability, etc.

Of particular importance for marketing is the two dimensional approach underlying the behavioral assumptions of marketing models. This approach was first proposed by Bass, 1969, who developed a diffusion model of the initial purchase (excluding replacement purchases) of "new" generic classes of consumer products. Bass proposes that the diffusion of a new consumer product is dependent on two factors, the coefficient of innovation and the coefficient of imitation, respectively. The innovation coefficient relates to the number of initial innovators buying a product, whereas the coefficient of imitation refers to the rest of the population assumed to imitate the behavior of the innovators and learn from their experience. Usually the group of innovators is characterized by high income, education and influence on decision making processes. Related to this two-dimensional behavioral foundation of the model is the distinction between external and internal influence diffusion models in marketing. This distinction relates to the communication channels through which information about a product is communicated (based on the work of Hamblin *et al.*, 1973).

External influence refers to information spreading vertically to potential adopters through mass media, advertising, etc., whereas internal influence refers to horizontal communication channels, i.e., interpersonal communication. It is generally argued (e.g., Gatignon and Robertson, 1986), that products that involve low consumer learning, are of low social relevance and are characterized by high

---

* This bias is represented by the fact that models generally deal with important (contrary to small incremental), investment intensive process innovations, the diffusion of which is not impeded by patents. The dominant role profitability plays as a causal driving force of the diffusion process relates only to process innovations, for which it represents the paramount adoption criteria used by firms.

** For an overview see Mahajan and Peterson, 1985, Mahajan and Wind, 1986, and Mahajan, Muller and Bass, 1990.

marketing/advertising efforts, diffuse via external influence communication, yielding a modified exponential diffusion pattern as first proposed by Coleman *et al.*, 1966. More complex and socially visible products requiring experimentation and observation of the experience by early adopters, and where non adoption tends to create some economic or social disadvantage (the type of products of most interest in this study), diffuse primarily through interpersonal communication channels with a resulting S-shaped diffusion pattern, usually represented by a logistic curve. The model proposed by Bass, 1969 (see also Bass, 1980), which can be considered as a "classic" in marketing models, combines both of these external and internal communication mechanisms. Bass, claims that the internal influence (i.e., interpersonal communication, as reflected in the coefficient of imitation) exerts a much larger influence on the successful diffusion of a new product, resulting in a logistic diffusion pattern.

Thus the rationale underlying technological diffusion and substitution models has also provided the basis for marketing models analyzing the diffusion of new consumer products. Whereas the relative role of influencing factors may be different from process innovation models, the formal analytical descriptions are very similar in both fields and allow similar conclusions about the regularity and driving forces of the innovation diffusion process. Still, it is somehow surprising that these two streams of diffusion research have never really converged or even interacted with each other. In particular, the relationship of marketing to technological substitution models has hardly been explored; this fact has been admitted by the representatives of the marketing models discipline (see e.g., Bass, 1986).

(4) The narrowness in the definition of influencing variables and their static nature

A final group of critical comments on diffusion models in economics deals with the narrow set of influencing variables included in the models, as well as their static nature, e.g., the profitability of an innovation is assumed to stay constant over the diffusion process. Finally, it has been argued that changes in the environment in which an innovation is embedded, for example, a changing competitive structure, is ignored by the models.

Probably the most extreme line of argument in this direction was followed by Gold *et al.*, 1970, who claim that innovation diffusion is affected by such a diversity of (changing) variables, that it is almost pointless to build general models of innovation diffusion. However, this appears to be an unnecessarily pessimistic view, as our main interest is in the general factors influencing diffusion against which special or random influences can be assessed further. The criticism of a narrow definition of influencing variables was certainly valid for the first generation of diffusion models, for instance, the models would not hold if the (relative) profitability of an innovation would be close to one. However, a large number of additional influencing variables have since been studied both in the technological literature (e.g., Sharif and Haq, 1979, who in addition to profitability and investment size, analyze factors like product quality, advertizing effort and price ratios), and the marketing literature (e.g., Gatignon and Robertson, 1986). Ayres, 1969, proposed an industry classification scheme, in which the driving variables of diffusion might be regrouped into (a) performance maximization, (b) sales maximization, and (c) cost minimization. Thus, the models have been considerably extended in the direction of a multidimensionality of influencing factors. Monocausality appears thus to be no longer a valid objection to the formulation and application of diffusion models both in the area of technological change and marketing.

Related to the missing multidimensionality of influencing factors is another critique on diffusion models: the static nature of the influencing variables considered. In reality these variables change continuously over time and interact with each other: e.g., technological performance will increase, prices go down, uncertainty in applications will be reduced, etc. This possible interaction has been neglected by a number of researches arguing that the profitability of adoption (and thus the the diffusion rate Δt) should decrease over time, resulting in an asymmetrical diffusion trajectory. However, it is rather the *expected* profitability, i.e., the vector of profitability and uncertainty about the future, that determines the diffusion rate. As profitability is reduced in due course in the diffusion process, uncertainty is reduced as well, so the resulting *vector* may also stay constant over time. In addition, there have been objections to the *ex post* nature of the measurement of the driving variables, like profitability or comparative advantage. Thus, the explicative power of these variables for the behavior of individuals (consumers or firms) would appear limited, as the decisions to adopt are rather the result of an *ex ante* (subjective) assessment of the comparative advantage of an innovation under conditions of uncertainty. This is contrary to the "objective" nature of the *ex post* assessed variables, assuming no radical uncertainty and quasi perfect information, usually taken for granted in traditional economic theory (see the papers contained in Dosi *et al.*, 1988, on a critique of assumptions in classical, equilibrium economics).

The dynamic nature of the driving forces influencing innovation diffusion applies not only to the innovation itself (e.g., through changing performance-price relationships, i.e., the well established learning curve effect) but has to be extended to the whole environment in which an innovation is embedded. Even an old technology, when challenged by competition may improve its technical/economic performance. This phenomenon has been termed the "sailing ship effect" (Ward, 1967, discussed further by Rosenberg, 1976), referring to the considerable improvements in the technology of sail ships (clippers) when challenged by the introduction of steamships. Whereas this effect can certainly not be denied, there exist some doubts about its impact on the diffusion of new technologies. Montroll, 1978, determined a time shift (of t_0) in the diffusion curve of steamships of 11 years, due to this effect which, in view of the time constants of steamship diffusion (75 years to diffuse from 10 to 90 percent market share in the USA), is certainly not dramatic. Similarly, by inspecting the empirical substitution curves presented in Section 3.1, one can observe a certain distortive influence of this effect. However, it does not seem to significantly affect the diffusion rate proper.

Although both old and new technologies are affected by changing performance, prices, and profitability over time, it appears to us that the most important element influencing diffusion is the perceived *relative* performance-price relationship between the old and the new technology. To go one step further, one might argue that it is not so much the perceived relative relationship at the time of adoption, but rather the subjectively assessed difference between the *ultimate* performance-price relationships of competing technologies, i.e., including the subjective assessment of likely future improvements along the learning curve.

The final criticism deals with the fact that diffusion models tend to ignore the *interaction* between the various influencing variables. In particular, the interaction between the supply and demand aspects of innovation diffusion has been noticed. This aspect is highlighted and incorporated into a diffusion model by Metcalfe, 1983 (see also Cameron and Metcalfe, 1987). Metcalfe is a prominent representative of the "equilibrium" diffusion modeling approach*. Here, diffusion is seen as a transition between equilibria levels, defined by changing economic attributes, e.g., costs, prices, etc., and a changing environment, e.g., differences in the market structure. Diffusion is not so much interpreted as a learning phenomenon, but as a result of the interaction of changes in the innovation and adoption environment, i.e., the interaction between suppliers and customers of an (product/process) innovation e.g., via licensing, price strategies and so on.

---

* For similar arguments and models see e.g., Peterka, 1977, and Silvennoinen and Väänänen, 1987.

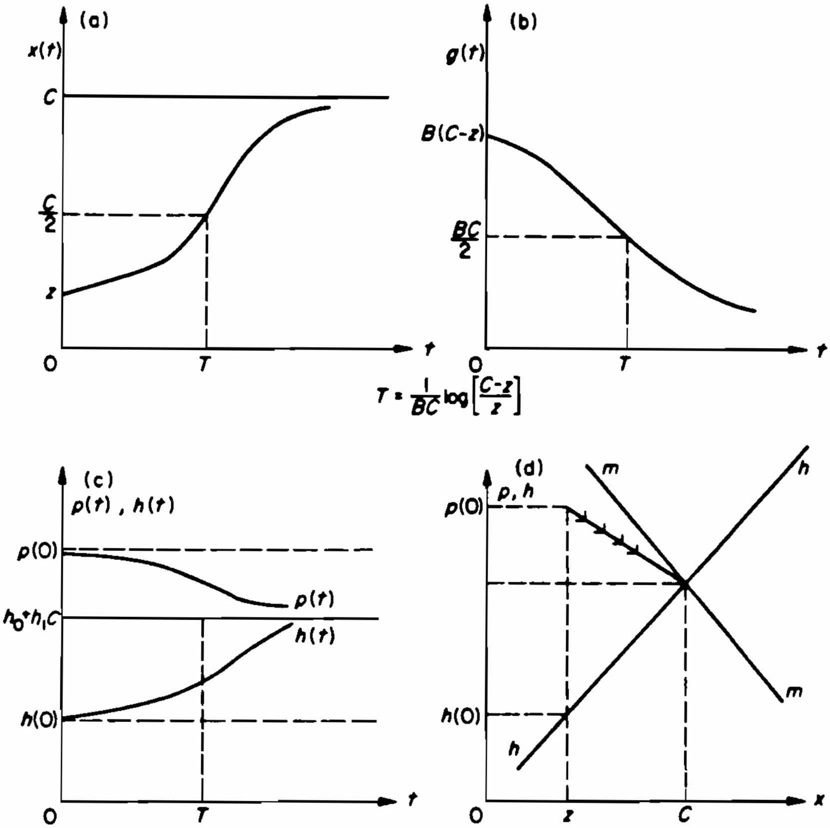


Figure 2.2.2. Innovation diffusion within an equilibrium framework. Innovation seen as impulse on the economy creating growth potential (from z to C), filled by the innovation. (Source: Metcalfe, 1983.)

Figure 2.2.2 presents a graphical representation of the most important relationships in the model proposed by Metcalfe. The term equilibrium is used in this model in two ways: first, supply and demand diffusion (i.e., the dynamics of demand growth and capacity expansion) are treated separately in that each has a different equilibrium asymptote, but are kept in balance (equilibrium) via price adjustment. Second, innovation diffusion is seen as the transition between two equilibria levels. All along this transition, relative prices, production costs, and the relative profitability of using and producing both the old and the new technology vary endogenously, with their variations being in turn driven by the substitution process. Without going into detail about the structure and mathematics of the model we summarize its main conclusions.

The substitution trajectory is determined by the long-run market share, the pre- and post-innovation market ratio and the (supply-demand) balanced output growth of the innovation, as well as the initial conditions that exist between the old and the new product/process. Yet, despite this complexity the substitution curve is reduced to a simple logistic [Figure 2.2.2(a)].

The innovation acts as an impulse in the economy creating a potential for growth, that is filled by the innovation. Adjustment to the new equilibrium level is achieved via the continuously retarding *combined* growth rate $g(t)$ [Figure 2.2.2(b)]. However, the price and costs of the new innovation change continuously over time producing a profitability squeeze, as shown in Figure 2.2.2(c). Finally, Figure 2.2.2(d) shows the equilibrium demand and supply curves. Note in particular that the price-output trajectory of the new innovation, i.e., arrowed line in Figure 2.2.2(d), does not lie either on the equilibrium demand or supply curves until the final equilibrium (end of diffusion) is reached. This points to the transient (non-equilibrium type) nature of the diffusion process, which we discuss more in Section 2.2.3 where we present the most recent diffusion models that are based on the principles of self-organization.

Models such as the one presented in Section 2.2.3 represent to some extent the end point of a long line of increasing complexity in the economic modeling of causal relationships underlying the diffusion process, a line of research initiated by the model of Mansfield. In the following Section we see that, despite the increasing complexity of innovation diffusion models, they are still consistent with the empirical regularities of diffusion identified and described by behavioral diffusion models. Thus, empirical regularities as identified in the subsequent chapters on infrastructure development and technological change in the transport sector, are consistent with the causalities of diffusion models formulated in behavioral sciences *and* economics. Structured evolutionary paths at the aggregate level emerge from (rather than contradict) *diversity* in behavior, technological characteristics, and economics at the micro-economic level or the level of the individual.

2.2.3. A self-organizing model of technology diffusion and structural change

This section* presents the underlying rationale and structure of a model in which the evolution of a particular system (i.e., the structural change induced by the diffusion of an innovation), is regulated by changing technological and behavioral *diversity*, *learning*, and *selection* mechanisms. These, together with their interrelated feedbacks, generate continuous adjustment processes to a changing techno-economic environment and result in ordered evolutionary paths at the macro (industry) level. The model proposed by Dosi *et al.*, 1986 (see also Silverberg *et al.*, 1988), appears particularly appealing as it reconciles both a complex set of economic driving variables and their interactions (and thus most of the host of hypotheses generated by first and second generation diffusion models discussed in the previous Section) at the micro level, i.e., the level of the firm, while still being consistent with the ordered evolutionary paths at the macro industry level, suggested by behavioral diffusion theory and empirical observations.

Characteristics of Technology and Industry Environments

Technology in a broad sense, i.e., including both process and product innovations is not a free good. Instead, it is characterized by varying degrees of *appropriability*, e.g., existence of patents, access to information, etc., *a priori uncertainty* about its future technical and economic characteristics and prospects, *cumulativeness*, in the patterns of innovation and capabilities to innovate (i.e., an analog to "learning by doing" as formulated by Arrow, 1962) and of *tacitness* of knowledge and expertise on which development and successful adoption of an innovation is contingent. Adoption decisions are characterized by particular *search and learning processes*, drawing on specific knowledge bases, containing both freely available information, like published scientific results, as well as internal and external skills.

As a result of the above, a fundamental characteristic of any industrial environment undergoing (technological) change is in fact the *diversity* of the economic, technological, and behavioral environment in which a particular innovation is embedded. Diversity in the economic environment implies that at any given point in time the economic population (be it firms or consumers) is in fact heterogeneous. Firms have different technological capabilities to innovate

---

* Based on Dosi *et al.*, 1986, and Silverberg *et al.*, 1988.

(consider for instance the different R&D expenditures and capabilities among companies), show different degrees of success in the development and adoption of innovations, and finally have different cost structures. Economic structure is the result of the differences in economic performance, the relative differences in innovation and adoption of innovations between individual firms, in search procedures, production techniques, combination of factor inputs, and finally the products of firms. The diversity between firms influences the rate and nature of the diffusion process in the following way: if the average *level* of technological capabilities in an industry is high, diffusion will proceed fast, if the *variance* of the distribution of capabilities between firms is high, one might expect *ceteris paribus* diffusion to proceed through competition rather than through learning/imitation.

Diversity between economic agents implies that any economic system undergoing change through innovation and diffusion, is in a disequilibrium situation in the neoclassical sense. "Better" and "worse" firms, with a different technological base (even to the extent of a redundancy of the technologies present in the market), etc., coexist. Decisions are affected by uncertainty about the technical and economic outcomes of the introduction of an innovation, because of interactions and interdependencies between firms with respect to prices, and technological and market competition, and so on. In fact, the outcome of a decision of any particular firm depends on the actions of other firms. Thus, it is difficult to reduce the behavioral diversity to a simple maximizing behavior underlying the traditional equilibrium type diffusion models.

The basic task of a self-organizing model consists thus in the representation of the feedback loops between the structure of an industry, the behavior of firms and the evolution of the industry in general. The coupled dynamics in the areas of technology, economic structure, and diverse behavior interact to produce relatively ordered evolutionary paths at the industry level, and in total at the level of the whole economy. The changing nature of the system in turn feeds back on the (technological) capabilities, incentives, constraints, and the behavior of economic agents.

Model Description

In the following we present an overview of a model proposed by Dosi, Orsenigo and Silverberg, 1986, with further discussion of the underlying economic rationale of some of the model assumptions and equations contained in Silverberg, 1984, and 1987. Dosi *et al.*, 1986 (p. 9) state:

> Innovation/imitation/diffusion [are]* represented as the process through which endogenously generated fluctuations of a system become 'autocatalytic' and, under certain conditions, progressively change the morphology [the structure] of the system itself. Diffusion of new products and production technologies is the outcome of evolutionary processes whereby the interactions between agents (the carriers of capabilities, technologies and behavior) induce changing incentives, selection mechanisms and learning processes. Innovation and diffusion processes are thus governed by (different combinations of) selection and learning mechanisms. Selection tends to increase the economic dominance of the firms which carry the innovation and penalize others, while learning spreads innovative/imitative capabilities throughout the (changing) set of potential adopters.

The balance/composition of these two different modes of diffusion contains technological, structural and behavioral components. More specifically, diffusion depends on (Dosi *et al., 1986, pp. 8-9):*

(a) the characteristics of each technology (sources of basic knowledge, degrees of appropriability and tacitness of innovation, complexity of research, production and products, existence and role of various forms of economies of scale, cumulativeness of technological learning, etc.),
(b) the degrees and forms of diversity between economic agents (including their levels of technological capabilities and variety of search procedures and behavioral rules), and
(c) the endogenous evolution of incentives, constraints and selection mechanisms (including the evolution of relative profitabilities of different technologies, firm sizes, cash flows and market shares).

The assumed industry situation for which the model has been developed, is characterized by a supplier dominated industry, which in turn purchases its investment goods from outside suppliers where no availability constraints are assumed. The market is characterized by strong price competition, although for the sake of simplicity no differentiation in the product according to different product qualities is assumed. Only one existing and one new technology are considered.

The diffusion of a new technology is represented by its incorporation into the capital stock of a firm, constrained by cash flow, i.e., no negative cash flow and self-financing are assumed in order to make the feedback between profitability and investment (diffusion) more straightforward, but at the same time also less realistic. In addition, scrapping of old capital stock is included in the model. Demand and supply are represented and linked by the variables orders, delivery

---

* Square brackets refer to alterations and emphasis added by the author.

delay (in case orders exceed productive capacity in the time interval), capacity utilization, and shipments. Obviously production is constrained by maximum capacity, and average and marginal costs are a decreasing function of the production level up to full capacity utilization. Total demand is assumed to be exogenous* (growing at an exponential rate), and demand (orders) for a particular firm are a function of competitiveness (price). Market shares of individual companies may therefore change over time due to disparities in their relative competitiveness. The competitiveness of a firm is defined in relative terms, i.e., the difference between its competitiveness and the average competitiveness for the industry.

Initially, the model consists of a system of equations in which a single best practice technology is available to all agents. It is assumed that the investment process (implicit in the payback method of the model) ensures that productivity gains due to technological advances (incremental innovations) are continually incorporated into the capital stock, even under different payback criteria used by different firms. Thus, the investment policy as represented in the model assures the diffusion of technical progress within the capital stock installed.

However, one has to go one step further: first by considering more than just one (old) best practice technology, although it too evolves in the direction of productivity improvements, and take into account the introduction of a** new technology that may, however, not be freely available to everybody. Second, although an innovation may initially have a lower productivity level than the best existing technology, it has a larger ultimate improvement potential. In this context, a new innovation might consist of improvements in the existing technology in terms of its economic and technical performance and capabilities. It could also constitute a "quantum leap" opening up entirely new performance dimensions, new markets, etc. which might not be apparent at the initial stage of introduction.

Firms making investment decisions have to subjectively weigh the improvement potential of existing versus new technology, i.e., the remaining improvement potential of their respective learning curves. In both cases, of course, the exact rate and ultimate potential for improvements are unknown. In addition, improvements in the technology are not only of a technical and/or economic nature, but involve also changing levels of expertise (scientific/engineering, qualification

---

* This model simplification appears problematic in view of the linkage between diffusion and market expansion, argued by Metcalfe, 1983, and discussed in the previous Section.

** The model deals just with the rather special case of a one to one competition between two process innovations producing the same type and quality of output, i.e., a fully standardized commodity. Quality differentiation, multiple competition and/or the possibility that a (radical) innovation opens completely new product lines/markets are not considered.

of workforce, etc.) available both internally and externally to the firm. Different strategies are therefore pursued by individual firms with respect to developing their knowledge base: either by in-house development or by waiting for the competitor to act and thus avoid (initially high) development and learning costs. These different strategies imply that an innovation might never be introduced if no one takes the initiative to develop it. On the other hand there is also the possibility that a firm might acquire a new technology free or at very moderate costs, after it has been developed, improved, and demonstrated as technically and economically feasible by some competitor. In fact, it can be shown by simulation runs that if the subjective assessment criteria (represented by an anticipation bonus about the future improvement of the technology) for a firm adopting an innovation and being the net "winner" are fixed uniformly for all firms in a subsequent simulation run, the innovation is not adopted at all, because no firm is ready to incur the costs of developing the technology and bring it to the commercialization stage.

Therefore the notion of a single "optimum time" for adoption has to be questioned due to the uncertainty and diversity of expectations about the future "trajectories" of existing and new technologies, and the actions and outcomes between the different agents thereof. Dosi *et al.*, 1986, go so far as to consider this diversity of expectations as not only unavoidable in any assessment for investment purposes, but almost as a prerequisite for the adoption of an innovation. They claim that this diversity is socially superior, despite the fact that it forces firms to profit or incur losses unevenly in the process of innovation adoption.

In order to model the dynamics of different technological trajectories the following assumptions are made. Two changing technological trajectories represented by maximum productivities are available. It is assumed that the second technology is ultimately always superior to the old one. The actual productivity realized by a firm is the result of the changing inherent productivity value of the technology and the changing specific skill levels of the firm. In the model firms are assumed to know only the *product* of the inherent technology efficiency and their respective efficiency in adoption and exploitation. They may (in fact *must*) make subjective assessments about the rate at which the new technology will achieve further productivity improvements as well as their internal rate of application efficiency (learning by doing). This can be improved both by internal learning, or by obtaining skills from outside, e.g., by hiring engineers, workers, etc., from competing firms. For reasons of simplicity the model assumes that the old technology is mature, i.e., its efficiency of application cannot be improved further. Modeling the evolution of the efficiency parameter of the new technology is quite simple as it is nothing more than the well established learning curve effect, where the rate of improvement (in use) is a function (power law) of the cumulative output (experience) associated with the technology.

The choice of a new technology will thus depend on a firm's assessment of its future learning curve potential. Investment decisions are awarded on the basis of an anticipative "bonus" with respect to its future prospects. The current realizable productivity of a technology is multiplied with this bonus and then compared to the best practice productivity of the old technology. The new technology will be adopted when its adjusted productivity is higher than for the old one and the new technology is either (1) cheaper per unit of capacity at the time of comparison, or (2) more expensive, but the investment difference can be compensated for by a reduction in production costs within the desired payback period.

Despite some simplifying assumptions, the model contains a clear economic rationale for the interaction and feedback mechanisms between economic agents. The complex structure of this model illustrates that there are many more mechanisms at work affecting the diffusion of innovations at the micro level than have been captured even in the most detailed diffusion models proposed in economics to date. It would be difficult indeed to infer the relative strategies, fate and performance of economic agents at the micro level, from a relatively small number of postulated driving variables, notwithstanding the explicative and predictive power of these simpler models at the macro level.

The following figures document some model simulation runs and illustrate how structured, evolutionary paths at the macro level evolve out of uncertainty, interdependence, and (antagonistic) competition at the micro level. The feedback mechanisms link the "whole" and the "parts" in order to demonstrate their mutual coevolution through a self-organizing mechanism. The characteristic diffusion pattern, formulated by traditional macro level diffusion models and observed empirically, emerges here as the outcome of the joint dynamics of technological behavior and economic activities and interactions at the micro level.

Figures 2.2.3 to 2.2.5 show the results of a simulation run in which* the pre-innovation equilibrium is disturbed through the

---

* The main assumptions for the simulation run are: The (new) technology 2 is potentially 100% more productive. Both technology 1 and 2 evolve at a rate of 4% per year as do nominal wages (production costs). The initially higher price of technology 2 decreases at a rate of 1% per year. All firms start with identical conditions (model parameters) except for their (subjective) "anticipation" bonus used to estimate the improvement potential of technology 2. This bonus ranges from 3.3 (i.e., the firm that introduces technology 2 first, evaluates its potential productivity a factor 3.3 higher than its present productivity) to 1 (i.e., the firm does not consider technology 2 to have any future productivity gain potential).

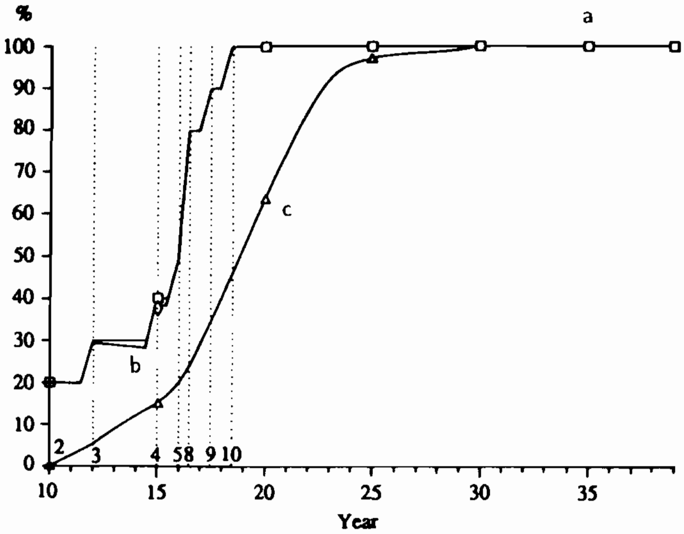


Figure 2.2.3. Diffusion curves of incorporating an innovation in the capital stock of an industry, (a) share of adopting firms in total number of firms, (b) market share of adopters, (c) percent capacity installed with new technology. (Source: Dosi *et al.*, 1986, and Silverberg *et al.*, 1988.)

availability and consequent diffusion of a new technology. The different strategies followed by different firms produce a complex set of phenomena and outcomes at the micro level, while resulting in a ordered path at the macro level. This is consistent with the empirical observations in the diffusion literature, as well as with the empirical regularities on transport technologies and infrastructure development identified in Chapters 3 and 4.

Figure 2.2.3 shows the aggregate macro level behavior of the system represented by three classical measures of diffusion: percentage of adopting firms (curve a), percentage of the market share held by those firms who adopted the new technology (curve b),* and finally (curve c) percentage of installed capacity, i.e., the result of inter- and intrafirm diffusion as well as of the changes in the relative sizes of firms. Dashed lines indicate the adoption dates of particular firms. All three measures display the classic S-shaped diffusion pattern.

---

* In this particular example the adopters (aggregated together) have not gained significant market shares over the non adopters. See, however, Figure 2.2.4 for the drastic differences in the market shares of the individual adopting firms.

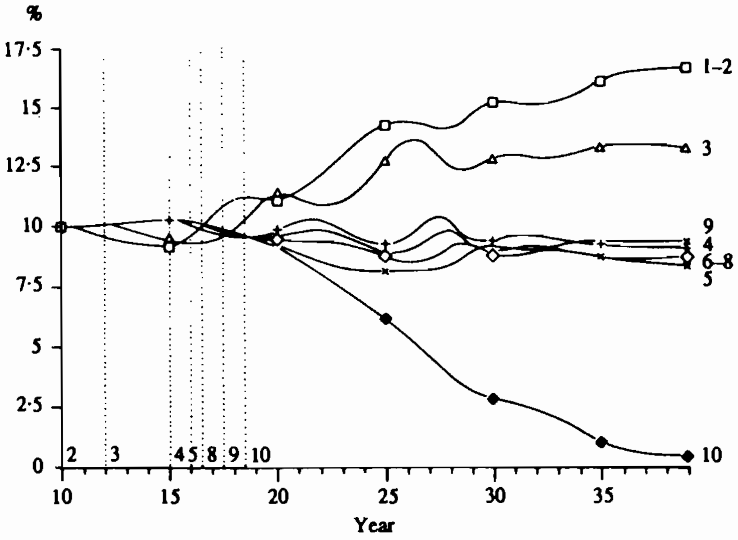


Figure 2.2.4. Market share of individual firms as a result of changes in competitiveness through adoption (or lagged or non-adoption) of an innovation by individual firms. (Source: Dosi *et al.*, 1986, and Silverberg *et al.*, 1988.)

Figures 2.2.4 and 2.2.5 illustrate the micro-economic "drama" going on underneath the smooth macro level surface. The relative success of the different firms in terms of market share (Figure 2.2.4) and in average productivity of the capital stock (Figure 2.2.5) is very different.

In terms of market share, early adopters of technology 2 do not perform as well as later adopters. This particular simulation outcome is (among other factors) the result of the assumed appropriability of technology 2, as reflected in the rate of internal learning. If the respective coefficient is set higher, the appropriability of technology 2 is accelerated, and the adopters become net benefiters. Particularly noteworthy is the market share of the "laggard" in adopting the technology 2: it is completely driven out of the market. This demonstrates the *pitfalls of missing the boat by not providing for an anticipation bonus in the productivity assessment of a new technology* (Dosi *et al.*, 1986, p. 33). The results show, that the relative payoffs of different adoption strategies depend partly on factors beyond the control of an individual firm, such as the adoption decisions of other firms, appropriability of a technology, etc. This illustrates the disequilibrium nature of the diffusion process, i.e., there is no individual optimal strategy independent of the strategies of one's competitors.

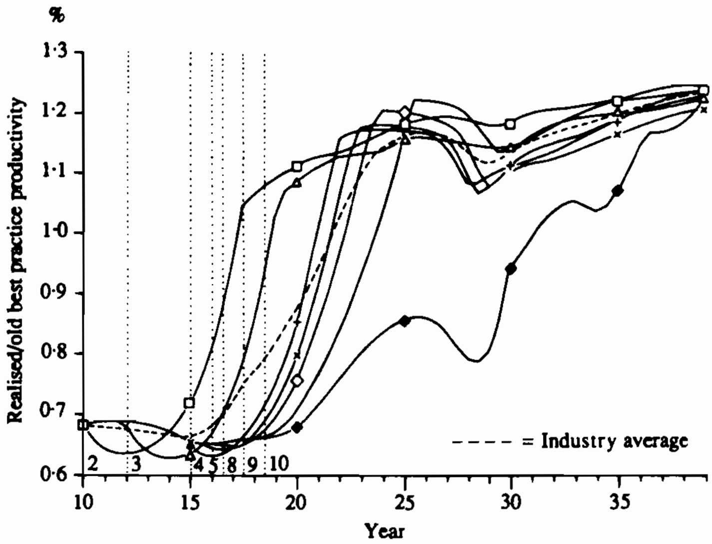


Figure 2.2.5. Realized average productivity of capital stock of different firms as well as industry average (dashed line) as a result of innovation adoptions. (Source: Dosi *et al.*, 1986, and Silverberg *et al.*, 1988.)

Figure 2.2.5 shows the productivity of the capital stock realized for each firm, as well as the industry average (dashed line in Figure 2.2.5). It shows in particular how successful adopters realize above average productivity levels, which contribute to their market share gains, whereas late adopters stay consistently below this average. This is a result of the differentials in the internal efficiency (skill) levels of firms. Whereas the innovation pioneers indeed build up their skill levels, later adopters benefit from this experience via external learning and eventually overtake the earliest adopters in terms of skill level and realized productivity. It should, however, be remembered that productivity is only one variable influencing competitiveness and thus market shares. The dynamics of the delivery delay incorporated into the model contribute equally to the changes in the market shares and the price margins realized.

Of course the outcome of the simulation run presented is the result of the particular technological, market, and behavior variables assumed. The main purpose here, is not to discuss the simulation result in terms of how well it represents reality, or in terms of "robustness" of innovation policies responding to the diversity in, and uncertainty about, the behavior of economic agents and the interdependencies of their strategies. Instead, the model demonstrates only the dynamic behavior of such an evolutionary self-organizing system. The main lesson to be learned is (the realization that) the dynamic interaction between the macro and micro level in such a system leads to the emergence of spatial and temporal patterns, which are *driven* rather than dissipated by micro level diversity. This becomes especially important when interpreting the empirical long-term regularities of diffusion and technological substitution. From this perspective, regularity in the evolutionary paths at the macro level is not a contradiction but rather a *consequence* of the diversity of technological expectations, designs, dynamic appropriability, and behavior of economic agents.

2.3. Spatial Diffusion Models

How do innovations diffuse spatially, and how do they affect spatial (geographical) structures? Innovations in the transport and communication sector, which are the subject of the present study, especially have important influences on spatial structures, as they alter the human space-time activity patterns. The geographical discipline of spatial innovation diffusion cannot be comprehensively treated within the present context (for an overview see, for instance, Morrill *et al.*, 1988). We briefly discuss, however, the conceptual and empirical contribution of Torsten Hägerstrand to the field of spatial innovation diffusion. This is not only because of his pioneering role both at the conceptual and formal level (computer based Monte-Carlo simulation techniques as early as the beginning of the 1950s), but his approach continues to be the most comprehensive formal conceptualization of spatial innovation diffusion to date.

The scale of spatial innovation diffusion is similar* to that of temporal diffusion models: *hierarchically decomposed.* As such, diffusion may be represented by the *fractal* decomposition of various space-time hierarchies. Different spatial scales, in a hierarchy of "cones of resolution" (Abler *et al.*, 1971) may be considered, ranging, e.g., from the international, to national, regional, and finally local level. This hierarchical decomposition of spatial scales is important for two reasons. First, social communication networks, which are – as in temporal diffusion models – at the core of spatial diffusion analysis, are different at the various levels of the spatial hierarchy. At the

---

* Complex systems are generally hierarchical, which in turn allows their abstraction in the form of mathematical models, as discussed by Simon, 1988. Consider for instance the diffusion of road infrastructure, automobiles, car components, etc. See also Chapter 5 on a proposed temporal hierarchy of diffusion processes in the transport area.

micro scale (local level), communication networks are primarily based on interpersonal communication linking individuals to each another. This communication network changes when considering higher level spatial aggregates from the regional to national, or even international level. Second, spatial innovation diffusion is characterized by a strong heterogeneity of diffusion patterns, as innovations do not spread uniformly in space, but proceed through a hierarchy of "innovation centers", in order to spread out from these centers to the hinterland.

In the following discussion of Hägerstrand's spatial diffusion model* (Hägerstrand, 1953, and 1967) we are primarily concerned with the lowest level of the spatial hierarchy, i.e., the local or subregional level. As a result, interpersonal communication plays a central role in the conceptualization of the diffusion process. This does not imply that interpersonal communication would be of lesser importance at higher spatial hierarchy levels (because the latter are in fact *aggregates* of lower level entities), but suggests that other forms of communication and information flows may also be at work at these higher levels.

Hägerstrand's conceptualization of the spatial innovation process is based on two principles. First, it consists of a model of the diffusion process itself and its empirical regularities. Diffusion is seen to proceed in a "wavelike fashion" (Hägerstrand, 1952) through interpersonal communication networks. Second, a probabilistic concept of communication behavior (hence the use of Monte-Carlo simulation techniques) is central in Hägerstrand's model, as represented in his "mean information field" (Hägerstrand, 1953, and 1967).

Spatial innovation diffusion in Hägerstrand's model consists of three stages.

Stage 1: Local concentrations of initial acceptances (initial agglomeration).
Stage 2: Radial dissemination outward from the initial agglomeration, accompanied by the rise of secondary agglomerations.
Stage 3: The condensing and saturation stage, in which filling-in occurs and diffusion eventually ceases.

Thus, three main characteristics define the spatial diffusion process:

(1) As in temporal diffusion models, an S-shaped pattern in the cumulative level of adoption.

---

* Hägerstrand formulated in fact three models, including further subvariants of Model III. These are not, however, treated here.

(2) A "hierarchy effect" by which diffusion proceeds from initial centers to secondary ones. Within the context of urban systems (central places) this implies a hierarchical spread of an innovation from larger to smaller centers, and from the centers to the surrounding hinterland.
(3) A "neighborhood effect" i.e., diffusion proceeds outwards from innovation centers, first "hitting" (Brown, 1981) nearby rather than far-away locations.

Diffusion in a spatial context is seen as changing the relative proportions of a population in a region between non adopters and adopters. The mechanism of this change remains to be discussed. In Hägerstrand's model, adoption is a function of the cumulative number of messages received about an innovation (e.g., its advantages, experience with it and so on). The exact amount of information required before adoption takes place, depends on (different) individual "resistance" levels, as a function of both individual and group characteristics. This resistance can both be in the form of "social resistance", should an innovation be inconsistent with given societal values, or "economic resistance", stemming from the relative attributes of an innovation, such as (high) costs, triability, etc., as discussed in Section 2.2 above.

The information flows between early adopters and the remaining population (i.e., how, and the number of messages an individual receives, before his resistance is overcome) define the time lag between initial adopters and followers. The flow of information is assumed to be a function of the *distance* between message emitters (adopters) and receivers (rest of the population). Later Hägerstrand conceptualized physical "barrier" effects like lakes or uninhabited areas, which, in addition to distance, act as further retarding effects on diffusion. These are formalized in the form of "zero" or "half" contact multiplicators on the (distance decaying) message flows.

The above concept of diffusion via cumulative message flows may, to some degree, be considered problematic. In approaching innovation diffusion through *information diffusion*, many other variables influencing information circulation are left aside, including an individual's different perception and interpretation of information, possible transformation of message contents, etc. In addition, social rather than geographical distance may ultimately determine the "efficiency" of communication channels (see in particular Pred, 1967, in his postscript of the English translation of Hägerstrand's *"Innovation Diffusion as a Spatial Process"* on these points). The concept of individual "resistance" has also to be considered as simplifying, as it does not differentiate between the attributes of *individuals* (social values, risk attitude, etc.) and the ones of the *innovation* proper (e.g., costs, relative advantages, complexity, etc.).

The last two elements of Hägerstrand's model deal with how the message flow between individuals can be described, and secondly, how the initial distribution of adopters is modeled. Communication is modeled in a probabilistic manner, where the *probability* that an adopter communicates with a non adopter is a negative exponential function of the intervening distance corrected for the existence of natural barriers. The initial distribution of early adopters or receivers of adoption messages is not deterministically fixed, but instead generated by drawing random numbers, hence the use of Monte-Carlo simulation techniques*.

At a conceptual level the use of Monte-Carlo simulation suggests that individual human behavior is usually so complex, e.g., in the decision to be (become) the first adopter(s) of an innovation, that their aggregate spatial expression is randomly determined within certain constraints, even though decisions behind individual behavior are not randomly motivated. Spatial diffusion in Hägerstrand's conceptualization (i.e., exactly to whom information is spread leading to adoption decisions) is defined by the neighborhood effect (operationalized by communication probabilities as a function of intervening distance) and random selection.

With respect to the formalization of the communication flows Hägerstrand defines a "mean information field" (MIF), in which the probability of communication is a negative function of distance between individuals. Space-wise, the MIF assigns different communication probabilities to the area surrounding an adopter, represented by grid cells, with the sum of all cell probabilities being one. Thus, the MIF is the area into which a message, whenever passed, will fall with a probability of one, with the fractional probabilities (of "hitting" a particular grid cell) decreasing away from the origin as a function of distance. The communication probabilities of the MIF are calibrated based on telephone communication data, but Hägerstrand also considered migration data as an indicator for particular spatial communication networks. The frequency of messages sent out is operationalized by considering discrete time steps in the simulation runs, with one message being passed per time step.

---

* This implies of course that there exist various outcomes of the spatial distribution pattern of adopters between the model runs. However, from an interpretative as well as conceptual viewpoint this is a significant advantage rather than a drawback of the model. Recall here the discussion of the evolutionary economic diffusion model in Section 2.2.3 above, where diversity in individual behavior results in ordered structured evolution paths at the macro level.

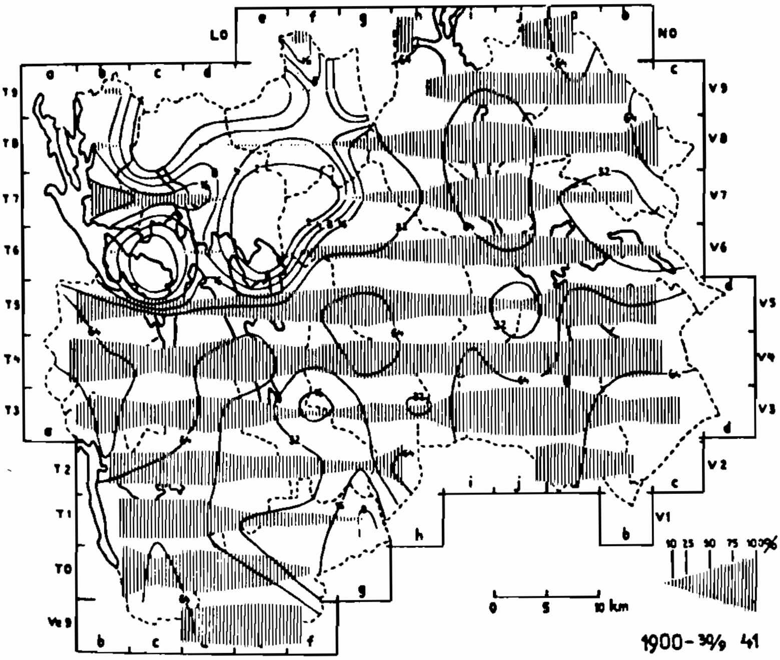
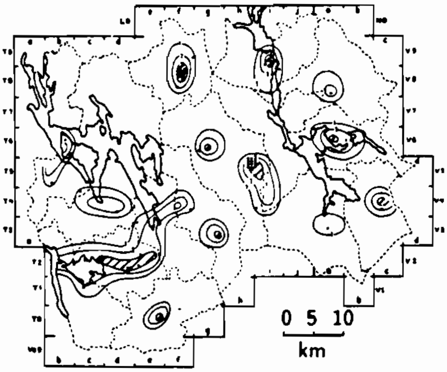


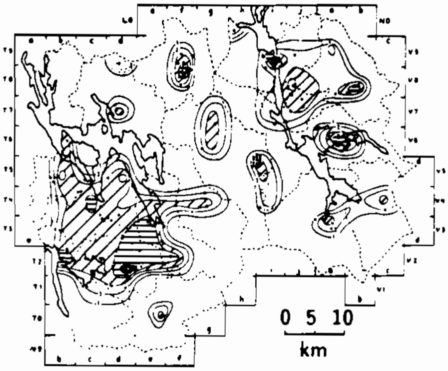
Figure 2.3.1. Spatial diffusion of tuberculosis control by farmers in two districts of southern Sweden. (Source: Hägerstrand, 1967.)

The model has been applied to study the diffusion of a number of innovations, including agricultural ones like tuberculosis (TB) control for cattle, financial (such as the diffusion of checking accounts), and transport and communication innovations (automobile and telephones) in two small districts (Kinda and Ydre) in the Swedish province of Östergötland.

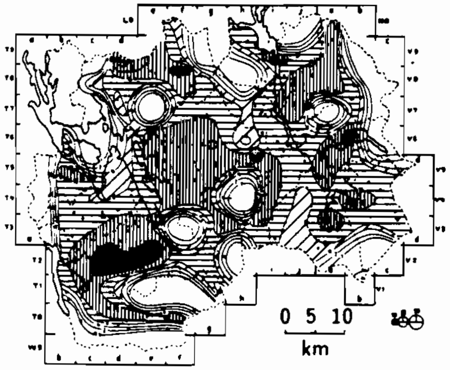
Figure 2.3.1 illustrates the "wavelike" fashion of the spatial diffusion of innovations on the basis of the adoption of TB-controls by farmers in the region. Compared to temporal diffusion, more complex patterns may be observed in *space*, as innovations spread out from a number of innovation centers and may overlap in between. Note that the model maps spatial diffusion patterns at the lowest level of the spatial hierarchy discussed above. If the "hierarchy effect" between central places is examined, e.g., in considering spatial innovation at a national level, the resulting pattern may become even more complex. This should be borne in mind when judging simplifying restrictions of higher aggregate spatial diffusion models, such as the one described in the following Section.



1920



1923



1933

Figure 2.3.2. Spatial diffusion of automobile ownership in a region of southern Sweden 1920, 1923, and 1933, in isolines of areas having more than 4, 8, 16 and 32 car owners. (Source: adapted from Hāgerstrand, 1967.)

As a conclusion to this Section, we present Hägerstrand's simulation results with respect to automobile ownership diffusion as presented in Figure 2.3.2. Particularly noticeable is the spread of the innovation out from a number of innovation centers. In terms of the diffusion speed and absolute diffusion levels reached, the spread of automobile ownership proceeds differently from the various centers as a result of spatial differences in communication networks and the "resistance" of the resident population to adoption.

Adoption levels are highest in innovation centers, i.e., in those areas where the automobile was first introduced. We show below (Section 3.3) that a similar observation can also be made for the diffusion of the automobile at the international level. This points to the self-similarity of the process of spatial diffusion along all spatial hierarchy levels. Also visible is the "wavelike" fashion of the diffusion: initially slow, followed by rapid growth, and finally very slow again as it reaches the farthest hinterlands of the innovation centers where adoption levels are lowest. This confirms the formal and functional equivalences of both the temporal and spatial dimension of innovation diffusion, pointing to their common conceptual basis: diffusion as a social learning phenomenon.

2.4. Diffusion Models Integrating Time and Space

Any social, economic or technological diffusion or substitution phenomenon occurs simultaneously in space and time. From the previous Sections, where we discussed temporal and spatial innovation diffusion models, the following regularities of the process emerged.

(1) Interpretation of technological change as a learning process, driven by a complex vector of communicational, social, and economic stimuli.
(2) Diversity in the behavior of economic agents, in available innovation designs, and uncertainty about the ultimate (economic and social) consequences of adoption decisions imply that an innovation is not accepted instantly by a social system, but spreads as a function of the experience gained.
(3) S-shaped patterns of diffusion/substitution result.
(4) There is a hierarchy effect in spatial diffusion.
(5) There is a neighborhood effect, i.e., diffusion spreads out from innovation (usually urban) centers, first reaching nearby rather than distant areas.

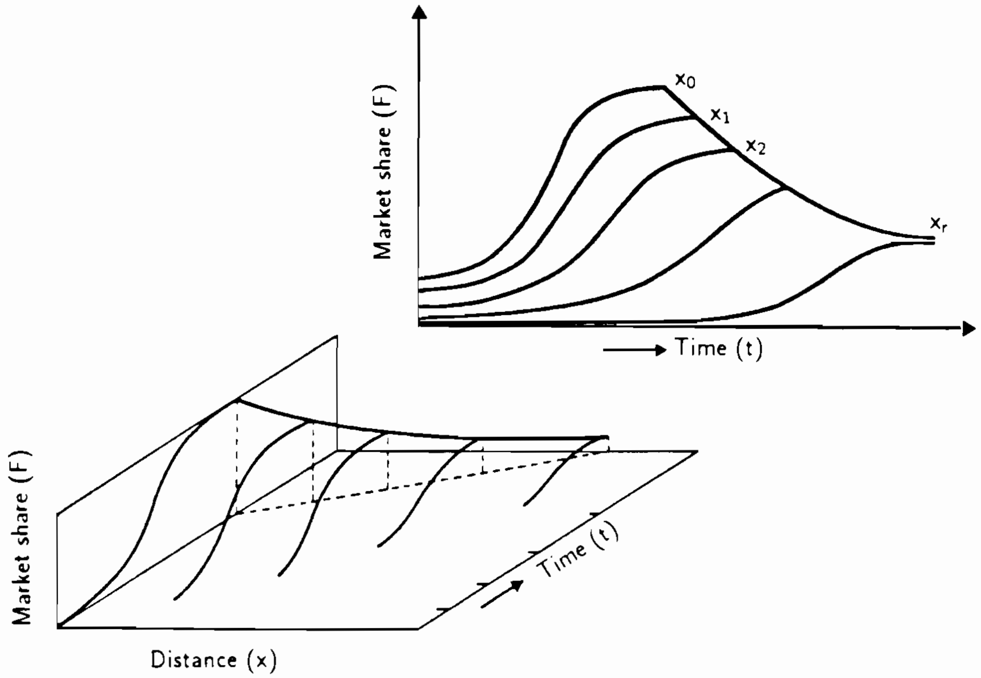


Figure 2.4.1. Substitution patterns as a function of time and space. (Source: Mahajan and Peterson, 1979, after Morrill, 1968.)

In the following we discuss an *intergrated* space/time diffusion model proposed by Mahajan and Peterson, 1979. The purpose of this model presentation is primarily illustrative. It is not considered to be an ultimate formulation of an integrated space/time diffusion model as it is affected by a number of limiting simplifying assumptions. However, its simplicity, and especially the fact that the model was tested successfully with empirical data, merits its presentation as an example of the class of integrated models dealing with spatial *and* temporal aspects of innovation diffusion.

Figure 2.4.1 presents a graphical representation of the integrated time-space model following the work of Morrill, 1968, and 1970, (see also Steinbach, 1980, for a discussion of the Morrill model). Diffusion (substitution) is considered to evolve over time following a standard diffusion model of the logistic type, with the time lag of introduction being a function of the distance from the innovation center. The upper limit of diffusion is assumed to decrease also as a function of distance from the innovation center, whereas the *diffusion speed* is assumed to be constant for all regions. Thus only K (the saturation level of diffusion) and a (the time shift parameter denoted in this study as t_0), but not the diffusion rate parameter b (Δt in our notation) of the standard logistic diffusion model are considered as functions of distance and time. Thus, denoting* the asymptote of the diffusion in the innovation center as K_1 and the distance from the innovation center to the peripheral regions 2,...,n with x, the asymptote in region x $\bar{F}(x)$ is defined by

$$\bar{F}(x) = K_1 - d\frac{x^2}{2} \tag{2.4.1}$$

where d is a constant, similar to the distance sensitivity parameter of standard gravity spatial interaction models. Assuming a constant diffusion rate b for all values of x (i.e., all regions), then

$$\frac{\partial F(x,t)}{\partial t} = b\,F(x,t)\,(\bar{F}(x) - F(x,t)) \tag{2.4.2}$$

and allowing the time shift parameter a to be a function of the distance x and using the Fisher-Pry variable t_0 (inflection point) to represent a with

$$F(x,t_0) = F_0(x) \tag{2.4.3}$$

as the initial value and

$$\frac{\partial F(0,t)}{\partial x} = 0 \tag{2.4.4}$$

as the boundary value, to ensure that the solution of equation (2.4.2) for the innovative region is independent of x. Without going into detail about the formal derivation of this model (described in Mahajan and Peterson, 1979), the final formulation of the space-time model is given by:

$$F(x,t) = \frac{(K - d\,x^2/2)}{1 + (((K - d\,x^2/2) - F_0(x))/F_0(x))\exp(-b(K - d\,x^2/2)(t - t_0))} \tag{2.4.5}$$

For x=0, equation (2.4.5) reduces to the standard logistic diffusion model with $\bar{F} = K$ and equation (2.4.4) also being satisfied. In their application of the model to the diffusion of farm tractor adoption in 25 states of the central farm producing region of the USA, Mahajan and

---

* We have retained the original notation of Mahajan and Peterson, 1979.

Peterson, 1979, conclude that the estimated regression coefficients of the model are statistically highly significant, and report a good fit between the adoption rates estimated via the model and the empirical data for the entire period.

Despite the fact that the model shows a certain elegance in its simplicity and that the empirical test did indeed generate impressive results, a number of serious model simplifications should be borne in mind. First, the measure of distance from the innovation center is only a crude approximation of the spatial aspects of diffusion. In particular, the "hierarchy" effect, where the innovation spreads through main, secondary, and tertiary innovation centers and then on to the hinterland (i.e., an innovation may spread out from more than just one center), is ignored in the model. Secondly, distance is in fact a surrogate measure: first incorporating one spatial aspect of diffusion proper, i.e., the spread of information and experience with an innovation from centers to hinterland areas, and second, reflecting the different characteristics of an innovation, e.g., its profitability of introduction, in different areas as reflected in the decreasing value of $\bar{F}$ with increasing distance from the innovation center.

Last and most important, the model assumes a constant value of b (Δt in our notation), i.e., diffusion in other areas proceeds at the same rate as in the region of origin. However, bearing in mind the learning process, we should expect a "catch-up" effect in peripheral areas, so that although the diffusion process started at a later date, it proceeds faster, shortening the overall diffusion time of late starters. We show empirical evidence supporting this hypothesis in Chapter 3.

Therefore, the above model should be interpreted mainly as an *illustrative* example, showing how an integrated formulation between the spatial and temporal aspects of diffusion can be achieved. A more comprehensive integrated model was formulated by Sonis, 1984, and 1986. This model integrates innovation diffusion and substitution models with multinomial Logit and Dogit individual choice and regional growth/decline models based on multiregional competition. This (the complexity of which does not allow a concise discussion here) is a particularly interesting model for its duality formulation between the choice behavior of individuals and the spread of competitive alternative innovations, seen as the driving force for urban/regional dynamics. As such, it supports the formulation of innovation diffusion as a structured, self-organizing dynamic process under behavioral and technological diversity.

2.5. Overview and Conclusion

In this methodological chapter we have presented a number of mathematical models to describe the *growth* and *structural change processes* from the perspective of the *interaction* between a (technological) system and its (social and economic) environment. Innovations in the transport sector, physically manifested in infrastructures and the technologies using them, are subject to competition and adaptation to changing social and economic *boundary conditions*. Adjustment processes take the form of structured evolutionary transition paths, which can be summarized by simple mathematical models.

We then explored at the macro level how such structured evolutionary paths can be rationalized from the perspective of sociology, behavioral sciences, economics, and geography. In a long line of increasing model complexity we reached the conclusion that the dynamics of technological and economic systems are best viewed as portraying the features of self-organizing systems. Structured evolutionary paths at the aggregate level emerge rather than dissipate due to *diversity* in behavior, technological characteristics, and economics at the micro-economic level or that of the individual.

Diffusion models serve as a useful tool for the systematization and organization of empirical data, in order to describe *the dynamics* of long-term development patterns. The models are also a prerequisite for comparative analysis between different countries where the absolute size of a particular market (the "niche" infrastructures grow into, or within which various technologies compete for market shares) is very different in terms of population, size, GNP, etc. Thus, the formal analysis framework used here allows us to take the self-referential qualities between different systems (countries) into account. The use of a formal descriptive framework also enables us to identify regularities and stable long-term trends in the transport sector. We show that the development of transport systems evolves along structured development paths both in terms of growth and decline. In addition, we document a regular pattern in the changing morphology of the transport system, consisting of a sequence of replacements of old by new forms to satisfy human (transport) needs.

The models presented in this chapter provide not only *consistency* for our comparative analysis, but equally an underlying *rationale* based on the findings from a number of research disciplines, ranging from social learning theory to evolutionary economics. The dynamics of change and the stable evolutionary patterns identified give an insight into the lead times for social acceptance, proof of economic viability, and possible rates of change in the competition between alternatives. The results suggest that these lead times are very long indeed, especially for large, pervasive technological systems. We also have to recognize explicit *discontinuities* in development, identified by the use of formal descriptive models. In Chapter 5 we show how these discontinuities in development can be interpreted from the perspective of *Wechsellagen* in long-term economic development (Kontradieff or Long Waves).

Having established a consistent descriptive picture, the next obvious question is what drives the system and why is there regularity in its evolution. At this point a final *caveat* is in order. Models aimed at describing driving forces and causality have become increasingly complex, in terms of the number of variables they consider and the multitude of interactions and feedback mechanisms, as shown in particular in Sections 2.2.2 and 2.2.3. This provides us with an even better understanding of the driving forces of technological change. Unfortunately, a similar complex and differentiated strategy to explain *why* we have come to the point we have with respect to our transport infrastructure cannot be followed.

Many of the key variables and system configurations responsible for the long-term development are, in fact, hidden from us in the historical past. Thus our modest efforts (in comparison with the complexity of an ultimate comprehensive causal model) to formulate some generalized propositions and aggregate driving forces responsible for the historical development patterns (as done in Chapters 4 and 5) should be seen as *indicative* rather than *comprehensive.*

We consider, however, that a *description* of the evolutionary pattern that shaped the transport system with its regularities and discontinuities provides useful insights. Contrary to the detailed descriptive portrait, we use a much wider brush to illustrate some of the underlying driving forces and causalities. As such the discussion is mainly *hinting* at some of the key variables that may be responsible for driving the system. The intention is to invite further exploration rather than provide an ultimate answer.

CHAPTER 3

Evolution of Infrastructures: Growth, Decline, and Technological Change

> *...the evolution of man's artificial aids to his effectors, by a process of "survival of the fittest" was recognized as early as 1757 by David Hume. We can now add, from the modern quantitative viewpoint, that this resemblance between the development of our inorganic accessories and that of organic populations extends also, in a number of instances, to the growth curves.*
>
> Alfred Lotka, *Elements of Physical Biology*, 1924
> (describing the growth of railways in the United States)

In this Chapter we try to give a quantitative overview of the rise and fall of different transport infrastructures in chronological order. The evolution of canals and ships, railways, roads, horses and cars, and finally, air transport is discussed. Illustrative examples of technological change and substitution for those technologies using particular infrastructure networks is also given.

The absence of detailed quantitative data led us to focus our attention on *long-distance* transport networks, thus, urban mass transport is not discussed in detail here.* Our analysis concentrates on a *quantitative* overview of the historical development patterns of industrial transport infrastructures and on a comparison of the dynamics of their growth and decline in different countries.

---

* For an excellent narrative and pictorial history of the London transport system the reader is referred to the two volumes of Barker and Robbins (1975).

Due to space limitations we refrain from supporting our quantitative analysis with the rich body of information assembled by transport historians on the numerous facets of infrastructure development and the multitude of – in retrospect even bizarre – technological options advocated at various periods in time. Here we refer only to some of the classical works that provide an excellent overview of the history of transportation: Sax, 1920; Voigt, 1965; various articles contained in the Oxford History of Technology (Singer *et al.*, 1972, 1975, and 1979); and the Murray History of Technology (Daumas, 1980). Finally, much useful information on the role of the transport system in the industrialization process is contained in the monumental work of Sombart, 1916. Detailed works on transport development in individual countries are Dyos and Aldcroft, 1969, for England; Taylor, 1962, for the United States of America (USA); and Toutain, 1967, for France. Where appropriate, further references on the history of individual transport systems will be made.

As the quote from Alfred Lotka at the beginning of this Chapter illustrates, a large number of scientists deserve credit for their pioneering contributions to the quantitative description of infrastructure development. Even at the risk of doing injustice to those who are not mentioned here, names like Alfred Lotka, Walter Isard, and the invaluable statistics compiled by Wladimir Woytinsky, Brian Mitchell, and Jean-Claude Toutain have to be mentioned. We feel the present study represents a new contribution in so far as it provides a systematic and common methodology based description of the development of *all* successive infrastructure systems for a number of countries, including also centrally planned economies. The methodological aspect becomes all the more important in view of the fact that infrastructural endowment levels cannot be compared internationally in absolute terms, because of differences in country size, population densities, spatial structures, etc. Our analysis shows however, that fruitful comparisons can be made of the *dynamics* of the processes of diffusion and technological change, once the self-referential qualities of different systems (countries) are adequately taken into account.

Thus, the *rates of change* and the timing of periods of growth, saturation, and senescence appear indeed to be comparable; however, we have to recognize explicitly *heterogeneity* in both absolute and relative diffusion (penetration) levels, in infrastructural endowment levels, and in the organizational embedding of similar infrastructure systems among different countries.

3.1. Canals and Ships

At the onset of the take-off of the industrial revolution in the mid-18th century, the transport system relied essentially on natural waterways and roads which had been developed to varying degrees in different countries. Both the natural and man-made infrastructures allowed for only modest amounts to be transported at high costs with low speed and quality of service. Arthur Young in 1770 (quoted in Daumas and Gille, 1980) spoke of "*infernal roads ... where you ran the risk of breaking your neck and legs*". On average transport speeds for passengers did not exceed about 3 kilometers/hour (km/hr), i.e., walking speed (Voigt, 1965). This was even true for France, where the highest quality transport infrastructure existed* as a result Colbert's efforts in early canal and road construction within the framework of a centralized mercantile policy. The transportation of goods by river or by roads was even slower (typically around 1.5 km/hr), and the costs were exceedingly high. In England it is reported that the cost of transporting grain by road from Carlisle to Newcastle was higher than the transport costs from Cape Hope to Newcastle (see Sax, 1920, and Voigt, 1965). Similarly, importing coal from England (over more than 1000 miles) was cheaper than transporting fuelwood over a distance of just 10 miles by road to the Atlantic cities of the early United States (Braudel, 1979).

Transport *systems*, in the contemporary sense of an *interconnected grid*, existed only in the most rudimentary forms. Consequently, longer distances were covered only by the most valuable "objects": human beings, information (mail), and high-value products (salt, spices, precious metals, textiles, luxury goods, etc.). The inadequacy of the transport infrastructure to carry large amounts of goods at modest costs to areas not endowed with natural infrastructures, like rivers or seaports (one of the most decisive locational factors for the foundation and growth of medieval cities), resulted in the repeated scarcity of local resources. Frequently the supply of food to areas where harvests had failed from other regions with abundant supplies was not possible, resulting in famines. Similarly, the so-called "fuelwood crisis" (see e.g., Sieferle, 1982) in 16th and 17th century England was not so much an "energy crisis" in that there was a lack of wood, but rather it was impossible to transport fuelwood and charcoal from areas with large supplies to the main centers of consumption (especially to the iron industries and cities). Braudel, 1979, and Sombart,

---

* Passenger transport speeds increased in France gradually, finally achieving, for the best connections, up to 12 km/hr by 1850 (Voigt, 1965).

1919, provide a comprehensive account of the inadequacy of the medieval transport system for economic development.

The first infrastructure system to overcome some of the deficiencies in the transport system of the pre-industrial age was the natural system of inland waterways; the waterways were greatly improved, and integrated into a high capacity transport *network* by the construction of canals.* Canal construction (especially after the invention of locks in 15th century Italy) did occur at a modest scale already before the onset of the industrial revolution. The first plans to link the Rhine-Main river system to the Danube, for example, date back to the period of Charlemagne, although a canal was finally built (with modest commercial success) only in the 19th century, and is at present an issue of lively public debate regarding its reconstruction.

However, the only countries where canal construction and river improvement occurred to any significant extent prior to the beginning of the 18th century were France and the Netherlands. The Dutch *treckvaarten* were used primarily for passenger transport and took considerable advantage of the topography of the Low Lands. Early French canals like the Canal de Briare (completed in 1642, linking the Loire to the Seine river to improve the connection to Paris) were, from the onset, conceived to integrate the inland navigation system into a national network for transporting goods and passengers. This was in contrast to early canal construction in England, which was primarily aimed at providing local access (e.g., from coal mines) to rivers and seaports.

Nevertheless, even in the case of France, the amount of construction going on was – whilst being impressive in view of the primitive contemporary construction technologies available – on a modest scale. The length of the French canal and improved inland waterway system increased from 156 km at the end to the 16th century to some 1,000 km at the end of the 18th century (Toutain, 1967). However, in England, the rapid early industrialization resulted for the first time in a "quantum leap" in the size and transport quality of the infrastructure network: the age of canals began.

England

The first attempts at improving inland navigation on rivers in England, by building locks and making cuts, resulted in 1,000 km of engineered rivers, in addition to more than 1,000 km of navigable

---

* The most comprehensive and detailed history of canal development is provided in the numerous books of Charles Hadfield, e.g., 1968 and 1986.

rivers by the mid-18th century. However, the poor spatial coverage of this transport system represented a major impediment in England's early industrial development. In fact, coal deposits located only a few kilometers away from rivers or seaports were practically unexploitable, due to the lack of transport possibilities. Early iron works had considerable logistics problems regarding the supply of basic materials (iron ore, charcoal, and later coal and coke) and for the delivery of final products (wrought iron).

It is not mere chance, that the first canal, which initialized the canal age in England, was built by the Duke of Bridgewater to transport coal from his Worsley mine to Manchester. The Bridgewater canal (1759–1761) became a symbol, visited and copied by many (especially continental) canal planners and developers. Generally canal construction in England was financed by private capital after being authorized by a parliamentary Canal Act. Widespread imitation of the successful Bridgewater canal resulted in a rapid increase in the length of canals as presented in Figure 3.1.1.* The total length of man-made waterways increased along a quite regular logistic growth pulse starting around 1760 and saturating around 1850 at a length of 5,630 km, or 96 percent of the (*ex post*) estimated saturation level. Hadfield's (1968) estimates of total canal length suggest a higher figure of around 6,570 km for England (including, however, Wales), but agrees with the saturation of the expansion in the length of the canal network by the 1850s.

Particularly noticeable in Figure 3.1.1 is the deviation from the long-term logistic growth trend around 1795. These years are characterized by feverish** canal construction and named by contemporary sources and canal historians as the period of "canal mania" with over 80 canal acts passed within a few years. This short-term deviation was, however, elastically absorbed in due course. The slowdown in canal construction after 1830 (when already over 90 percent of the ultimate maximum canal length had been constructed) is generally

---

* In the following Figures the appropriate parameters of a (logistic) diffusion model (saturation level K, inflection point t_0, and time constant Δt, i.e., time required to grow from 10 to 90 percent of the ultimate saturation level (or of market share) will be reported. **Note in particular the linear transforms $F/(1-F)$ on the logarithmic scale of the logistic diffusion curves, with $F=y/K$.** As the statistical uncertainty involved with parameter estimation is not significant when describing growth/substitution processes which are completed, no estimates of goodness of fit or parameter uncertainty for K will be presented. For diffusion/substitution processes, which are still ongoing, parameter uncertainty and R^2 values will be reported.

** One should bear in mind however, that even in peak years investment in canals and inland navigation represented not more than one third of total infrastructure investment (including turnpikes, parish roads, and bridges), and even all transport infrastructure investment combined together, never significantly exceeded one percent of the national product of the time, based on estimates of Ginarlis and Pollard, 1988. Still, even when we do not neglect the role of early road transport, we have to acknowledge the "quantum leap" in transport volume and quality resulting from canal development.

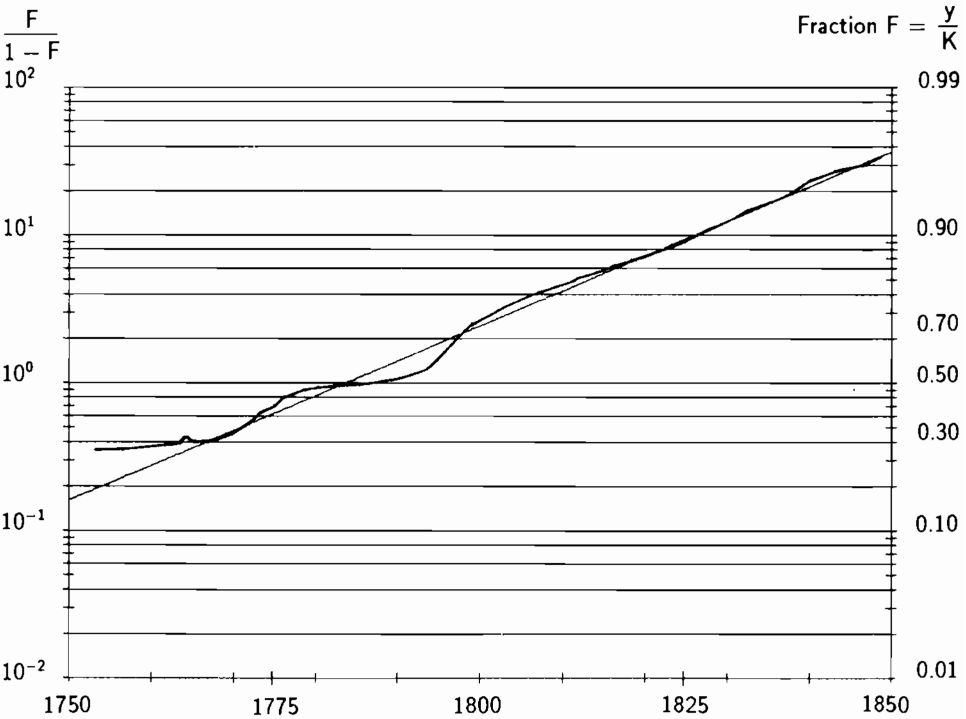


Figure 3.1.1. Growth of canal length in England [K 5,880 km, t_0 1784, Δt 84 years]. (Data source: Ginarlis and Pollard, 1988.)

attributed to increasing competition from the railways (e.g., Hadfield, 1986). However, the situation appears to be more complex, as suggested by Figure 3.1.1. The expansion dynamics of the canal network in England do not appear to be directly affected by the appearance of the railways in the 1820s. Based on our simple model, one could have predicted saturation in the expansion of the network as early as 1800, i.e., significantly before the advent of railways. In addition, with the introduction of the railways no significant deviation from the long-term logistic growth trend in canal network expansion can be observed (apart from the fact, that the growth process stops at 96 percent of its estimated ultimate potential). The main effect of railways on canals thus appears to be *after* the completion of the canal expansion pulse. This is reflected in the fact that no second growth pulse of canal network expansion was initiated, and also that railways started to compete with canals, resulting either in the closure of canals or their take-over by railway companies.

Before turning to a discussion of what happened to canals after their expansion saturated around the 1850s, let us acknowledge both the engineering achievements and economic impacts of canals. The construction of over 4,000 km of canals in the period 1760 to 1850 is in itself a formidable engineering achievement of early industrialization. Moreover, there existed an elaborate support infrastructure, e.g., locks, water reservoirs, etc., for water control and for overcoming height differences. The most spectacular engineering achievements were possibly the tunnels and aqueducts constructed for canals. Hadfield, 1986, estimates that by 1850 some 60 km of canal tunnels (the Standedge tunnel being nearly 5 km long) and a large number of canal aqueducts had been erected, e.g., like the impressive iron trough at Pontcysyllte – constructed in 1805 with a maximum height of 38 meters above the valley floor and over 300 meters long. The economic impacts of canals were also impressive in lowering transport costs and facilitating the growth of new industrial centers. For example, after the completion of the Bridgewater canal, the number of workers in the pottery industry in the area increased from 7,000 to nearly 20,000 (Voigt, 1965). The relatively reliable* and large-scale transportation of coal, industrial raw materials, food, and manufactured goods made possible by canals, reduced inland transport costs significantly. Canal construction and operation was also good business at that time: dividends of 70 percent annually or even over 150 percent in peak years are reported (Sax, 1920, and Duckham, 1983). However, with the expansion of canals approaching saturation, and with increasing (price) competition from the rapidly growing railways, profitability declined sharply, as discussed in more detail in the following Section on railways.

Other countries

The successful English model of canal development, especially the canal mania of the end of the 18th century was also adopted in continental Europe and in the USA, albeit with a considerable time lag. A similar growth pulse in Europe, comparable to the English canal development, started only after the Napoleonic wars. Still, despite this time lag, growth proceeded much faster and as a result, the expansion of the canal networks reached saturation in all industrializing countries, where quantitative data allows such an assessment, over a relatively short time period around 1860. This synchronization of saturation is worth noting, since all countries started with different initial conditions and expanded their networks with different growth rates to different ultimate maximum levels. We will see in due course,

---

* The main limiting factors were water shortages and the cold winter climate (ice). For estimates of the impact of climatic variations on canal operation see Freeman, 1980.

that an identical pattern can also be observed for other infrastructures and technology expansion pulses in later time periods.

It should be noted that canal construction did not come to a standstill even after the network had reached its maximum size and subsequently entered a period of decline. Examples are the Manchester Ship Canal completed in 1894 in England (built to cut high tariffs charged by railways and the Bridgewater canal which was controlled by railway companies after 1872), or the Canal de l'Est in France completed 1880–1882 (in order to compensate for the Lorraine waterways lost to Germany after the 1871 war). However, these canal additions could not compensate for the length of canal infrastructure being closed down after the 1860s as a result of the success of the railways.

Thus, canal construction after the 1860s does not contradict our conclusion about the synchronous saturation of the expansion of canal network size in a number of countries. With the exception of Germany and Russia (discussed below) canals everywhere reached their maximum network size around the 1860s, and entered a phase of decline or at best of stagnation thereafter.

France and USA

We have discussed above, that France took a pioneering role in very early canal construction, especially in the 17th century. By 1760 (the beginning of the English canal expansion) France had already developed over 1,000 km of engineered rivers and canals. However, as a result of the French Revolution, canal construction and maintenance came to a standstill. Despite the efforts of Napoleon, the network grew only modestly, i.e., by some 200 km between pre-1800 and 1813. Only after the Napoleonic era can observe the take-off of a similar (logistic) canal network expansion pulse, as in the case of England, with an ultimate saturation level of around 4,825 km being reached by 1867 (99 percent of the estimated K) as shown in Figure 3.1.2 (data from Toutain, 1967, quoting contemporary sources). The French canal expansion of approximately 3,570 km in the period 1820 to 1867 was based on a planning effort on a national scale, with resulting laws being passed in 1821 and 1822: *canales undique versum effosi* (Hadfield, 1986). Although the English canal system served as a general model, the dimensions of the canal locks were more modest than in the original plans, in order to speed up construction and lower costs. Financing was, in contrast to England, a combination of private capital and government financing. For the period after 1870, we have only scarce quantitative data on the length of the canal system proper. This evidence suggests, that the length of the canal network in France

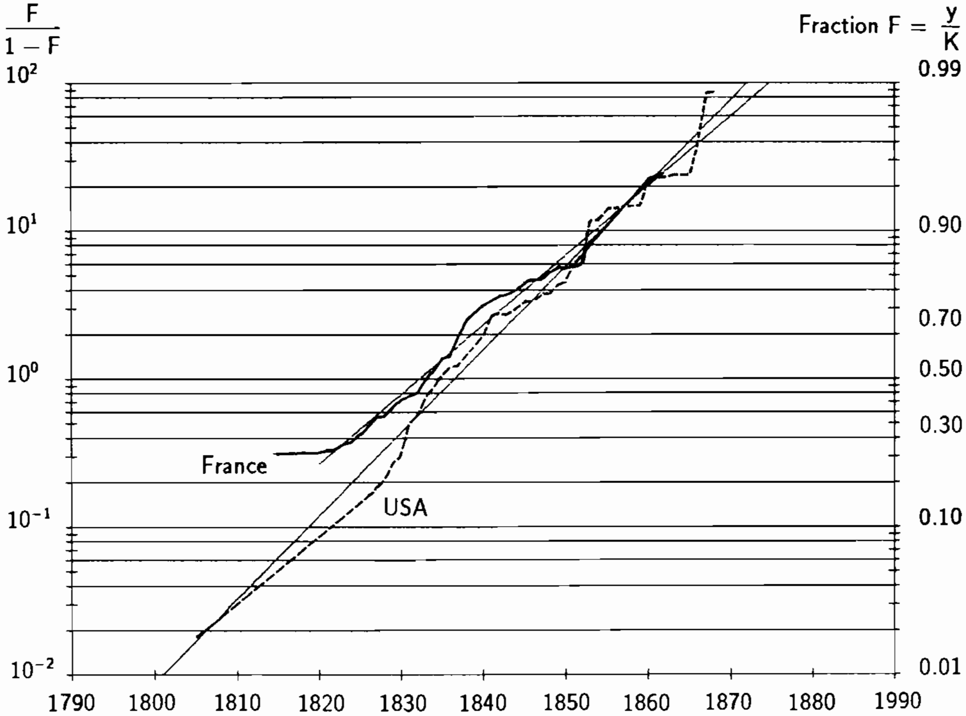


Figure 3.1.2. Growth pulses in canal expansion for France [*K* 4,825 km, t_0 1832, Δt 41 years] and USA [*K* 7,010 km, t_0 1835, Δt 31 years]. (Data sources: Toutain, 1967, and Isard, 1942.)

did not decrease significantly after the saturation of the expansion phase by 1870, but remained rather constant until the 1920s (Voigt, 1965, and Woytinsky, 1927). According to the most recent estimates, the French navigable waterway system (canals plus rivers) extends over 8,500 km, out of which 6,400 km are actually used for inland navigation (Ann. Stat., 1987). Thus, the extent of canals in use appears to have considerably decreased since the 1920s, with some of them finding new uses for recreational purposes.

The expansion of the canal network in the USA is* similar to that of France, except that it started practically from zero. According to Isard, 1942, only around 50 km of canals had been constructed by 1794. In due course, however, the length of the canal network increased along the growth pulse shown in Figure 3.1.2 to an estimated saturation level of some 7,010 km, with the maximum length being reached by 1861 at 6,860 km or 98 percent of our estimated *K*. The diffusion process in the USA, the country to start

---

* More details on canal development in the USA can be found in Goodrich, 1960; Goodrich *et al.*, 1975; Poor, 1860; and Taylor, 1962.

last in our sample, has a Δt of about 30 years and thus proceeds faster than both England and France. Particularly noteworthy is the synchronous saturation of canal expansion in France and the USA as reported in Figure 3.1.2. As a result of a "catch-up" effect, canal expansion in the USA reached saturation at a date very close to that of England and France, where canal construction began earlier, but where growth proceeded slower.

The financing (see Myers, 1970) of canal expansion in the USA was achieved partly by private capital but often with a speculative element. For instance, shares for a newly founded canal company could be easily obtained; collecting the funds from investors was a more difficult matter. Very often public funds (primarily from federal states, or cities) were also required. Until 1830, federal funds were available for transport infrastructure development in the USA, primarily for canals. Thereafter, local governments found more and more ingenious ways to raise funds for canal development (including a lottery by the state of Georgia), especially after the commercial success of the New York Erie canal (1825) became apparent. These financial engagements (together* with the funding of railway construction after 1830) brought a number of US federal states close to bankruptcy.

After the saturation of canal expansion in 1870, the total length of US canals decreased very rapidly under competition from railways. Perhaps the best indicator of the decrease in importance of canals is that no detailed statistics on the length of the canal network after 1870 are readily available. Voigt, 1965, reports that by the end of the 19th century over 4,000 km of the canal network in the USA (maximum length: 6,860 km) had been closed down. Canals in the USA (and to a lesser extent England), thus provide a first example of an infrastructure decay process after the zenith of its maximum expansion, a situation we will return to when discussing the railway network in these countries.

At this point we would like to stress another important aspect of the synchronization of the saturation of canal expansion. As the quantitative results presented above indicate, maximum canal network size occurred at different absolute and relative (in terms of per km^2 country area or per capita) levels. It seems therefore, that only the *growth dynamics* are comparable between different countries, and comparisons of diffusion levels between countries (in a kind of normative approach) is not possible. We suggest thus, that the diffusion of

---

* Out of the total US state debt of 175 million US dollars in 1838, canal shares accounted for 60 million, railroads for 42 million, and turnpikes and miscellaneous purposes for 20 million US dollars. Thus canals accounted for one third of state debt and transport infrastructure for 70 percent of US state debt (Myers, 1970).

canal networks (and we will see later, that the same applies also for the diffusion of railways and cars) is comparable in terms of the dynamics over time, whereas the resulting different saturation levels are caused by the diverse boundary conditions in which an innovation (canals in this case) is embedded in various countries. Such boundary conditions reflect differences in geography, endowment of natural infrastructures (like rivers), economic and spatial structure, and so on.

Canal network expansion after 1870

Although detailed quantitative statistics on canal expansion in other countries are lacking, there exists evidence that at least two countries are exceptions to the rule for the zenith of canal network length prior to 1870. These are Germany and Russia (USSR).

Sombart, 1916, reports that by 1836 some 650 km of canals had been built in Germany. Borchard, 1968, reports an estimate of 870 km for 1870. The canal network in Germany therefore appears significantly smaller than in England, France or the USA. However, especially after the creation of the German Reich in 1871, significant canal construction occurred. For 1913, Woytinsky, 1927, reports the canal length to range from 1,711 to 2,114 km. Still, the available quantitative data is too sparse to allow a formal analysis as for England, France, and the USA. It is not surprising that the expansion of canals continued after 1870 in Germany, as a consequence of the integration of Germany into a national state similar to France and England. As examples of such canal construction we can cite the Kiel (Kaiser Wilhelm or Nord-Ostsee) canal completed in 1895 (enlarged between 1909–1914) or the Dortmund-Ems canal with its 269 km completed in 1899.

The data situation with respect to Tzarist Russia and later, the USSR, is to some extent better, due to the research of French, 1983, and available official USSR statistics. The vastness of the country and the extensive length of its natural waterways, currently around 140,000 km, (Narod. Khoz., 1986) made investments into canals to link the various extensive river systems a necessary condition for the development of an early transport infrastructure. As in many other areas, the early pioneer in canal construction was Peter the Great, whose "canal mania" became even anecdotal. The most outstanding project was his attempt to link the Volga river system with the Baltic; this was of special importance after the foundation of Petersburg. The Tvertsa and Tsna canals (completed in 1708 and 1722, respectively) are early examples of successfully completed canals in Russia. The first "quantum leap" in the canal network came with the

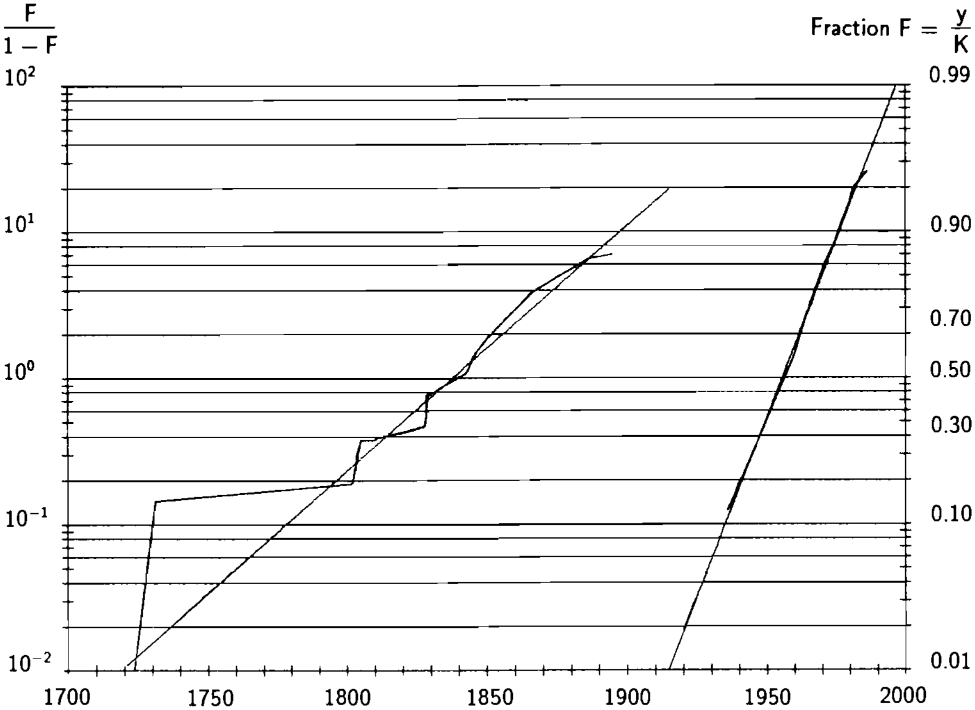


Figure 3.1.3. Growth of canal length in Tzarist Russia [K 904 km, t_0 1838, Δt 114 years] and in the USSR [K 21,710 km, t_0 1955, Δt 39 years]. (Data source: French, 1983, and Narod. Khoz., 1982, and 1986.)

completion of the 110 km long Emperor Peter the Great canal in 1731 (Figure 3.1.3). However, this initial burst in canal construction was not followed by further expansion, so that the long-term expansion trajectory (as represented by the logistic model in Figure 3.1.3) was resumed only at the beginning of the 19th century when a canal age similar to that in Europe started (Figure 3.1.3, based on the account of canal construction in French, 1983).

It becomes apparent from Figure 3.1.3, that the canal expansion growth pulse of Tzarist Russia proceeds substantially slower than in other countries (Δt of over 100 years) and extends beyond 1870, with important canals like the Emperor Alexander III or the Ob-Yenissey canal system being completed just before the end of the 19th century, at which point the length of the canal system had reached around 90 percent of the estimated saturation level of 900 km. This canal network extension does not appear large in comparison with other countries. However, one has to keep in mind the *complementary* and *integrative* nature of the canal network, linking the vast natural waterway system of Russia.

Contrary to what occurred in other countries, where the canal network decreased in length or at best stagnated after the completion of the 19th century expansion phase (being improved then only in terms of quality and capacity), we observe that the USSR followed an ambitious canal expansion program in the 20th century, extending the canal network from some 900 km* in the Tzarist period to close to 21,000 km at present. This second expansion appears to be perfectly described by the logistic growth pulse depicted in Figure 3.1.3, which is especially noteworthy for infrastructure development in a planned economy. Canal expansion in the USSR at present is almost saturated, and no noteworthy further expansion would be expected on the basis of Figure 3.1.3, a situation which becomes the more plausible, when considering the increasing (environmental) opposition to large-scale river improvement and canal construction inside the USSR. Despite the ambitious program of canal development, canals represent currently less than 2 percent of the length of physical transport infrastructures of the USSR (discussed in more detail in Chapter 4).

Technological change in the ship fleet

The decisive speed advantage of railways as a result of steam power, resulting in subsequent stagnation and decline of inland waterway infrastructures, led quickly to the adoption of the steam engine as a ship propulsion system. Whereas steamships on inland waterways** could not significantly improve the competitive position of inland navigation against railways (notwithstanding many successful local applications), they started successfully to compete with sail ships for long-distance (sea) transport.

In fact, the diffusion of steam propulsion in sea vessels, or rather the substitution of sail ships, was a relatively slow process. Challenged by the appearance of steamships, sail ship technology was

---

* This starting level is subtracted as an intercept in the calculation of the second growth pulse of USSR canal expansion for Figure 3.1.3. As the second growth pulse in Figure 3.1.3 is not an *ex post* description but represents a forecast, statistical measures of goodness of fit are presented: Assuming a 5 percent data error the estimated saturation level K (21,710 km including the intercept) is with 90 percent probability within 21.0–22.4 thousand km. R^2 of the estimated growth pulse is 0.998. The fact that we describe canal expansion in Russia and later the USSR by a discontinuous two phase diffusion model might seem at first sight arbitrary. However, we rationalize this discontinuous model by the inherent (and different) dynamics of the first expansion pulse as early as 1850, which allows us to conclude *ex post*, that its saturation will be by the beginning of the 20th century.

** Early examples of steamships for inland navigation are according to Hadfield, 1986, the *Margery* servicing the Thames, starting 1815 (renamed *Elise* and servicing the Seine after 1816). In Germany steam boats appeared on the Weser by 1817 and as an example of technological pioneering in Austria we would like to mention the 60 hp steamship *Franz I* introduced on the Danube by 1830 by the *Erste K.K. priv. Donau Dampfschiffahrt Gesellschaft* (DDSG), a company still operating (though heavily subsidized) on the Danube.

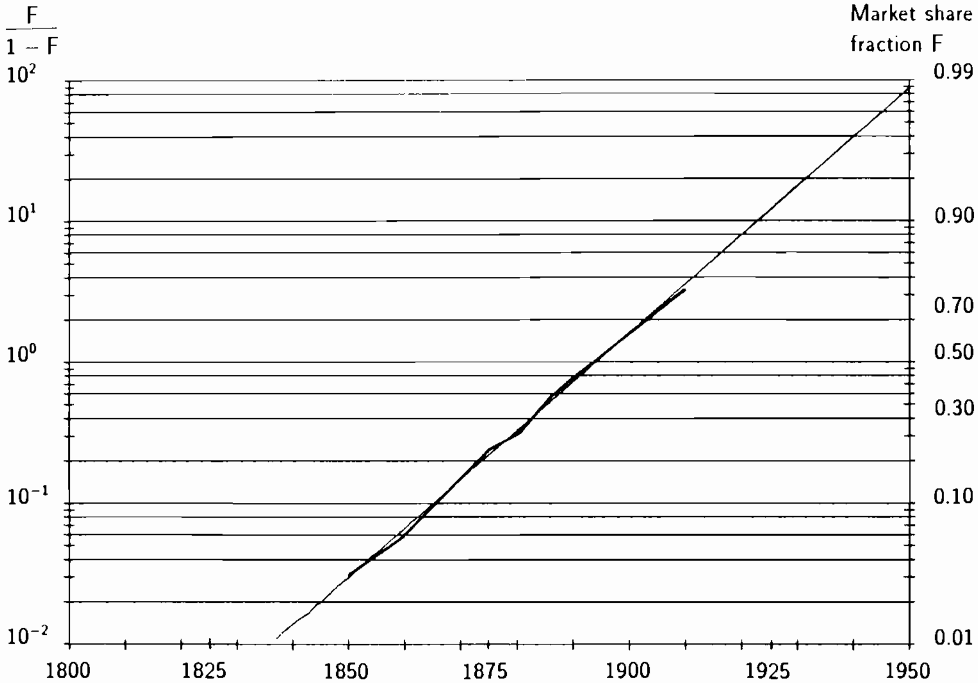


Figure 3.1.4. Share of steam- (and later also motor) ships in the gross tonnage of the world commercial ship fleet [t_0 1894, Δt 55 years]. (Data source: Woytinsky, 1927.)

improved to such an extent that even a new clipper age began.* Sail ship construction continued and the gross tonnage of sail ships increased even after the appearance of steamships. In analyzing, however, the relative** importance of the two propulsion systems in total gross tonnage, we can observe a regular substitution pattern along a logistic trajectory, reflecting the long-term comparative advantage of steamships. This is reported in Figure 3.1.4 for the total world commercial seaship tonnage. The excellent fit of the substitution model to the empirical data suggests, that the "sailing ship" effect is not discernible in the global substitution pattern.

As Figure 3.1.4 illustrates, the replacement of sail ships in the gross tonnage of the world's merchant fleet was rather a long process (Δt of 55 years). Although steamships offered considerable

---

* This so-called "sailing ship effect" (Ward, 1967, and Rosenberg, 1976) was already discussed in Section 2.2.2 above. As will be shown in the subsequent examples, this effect, while of undeniable importance in terms of the technology development of sail ships (design and performance improvements), did however *not* significantly affect the substitution pattern of sail by steamships.

** For substitution patterns, measuring *relative* market shares, K (being 100 percent) will not be reported in the subsequent figures.

advantages in terms of reliability (no dependence on wind conditions) and speed, sail ships continued to stay on the market for some 100 years after the introduction of steamships. This was made possible both by technological improvements as well as their successive movement into lower value market niches. Sailships progressively gave up transporting passengers and high-value goods and concentrated on goods where speed was not such a decisive economic criterion. Ironically, one of these lower value goods was coal for steamships, which continued to be an important market segment for sail ships well into the end of their life cycle as observed by Marchetti, 1987.

Available data does not permit the analysis of the replacement of steamships by motor ships for a large number of countries, as reported by Nakicenovic, 1984, for England (Figure 3.1.5 below). From this figure it becomes apparent that steamships reached their maximum market share by the 1920s, before starting to decline due to competition from motor ships. On the other hand, the completion of the replacement process of sail ships (below one percent of the market share) is estimated in Figure 3.1.4 to occur somewhat later, by 1940.

For the purpose of an international comparison of the dynamics of technological substitution in ship propulsion, it will suffice to analyze the most fundamental transformation process in civil shipping: the replacement of renewable (wind) by fossil energy (coal and later also oil) as a source of the required propulsion energy. One has to bear in mind, however, that the saturation dates reported in Table 3.1.1 refer to the completion of sail ship replacement rather than a correct measurement of the saturation date of steamships, due to the reasons outlined above. However, for our argument of the *synchronization* of development, especially of the saturation phases of technological substitution processes in different countries, this simplification is of minor importance. In Table 3.1.1 we have regrouped the parameters of a logistic substitution model, calculated from the empirical replacement data of sail ships in gross commercial tonnage for various countries (as reported in Figure 3.1.4 for the world total). Parameters for the United Kingdom (UK) are derived from Figure 3.1.5 (Nakicenovic, 1984) and for the USA from Montroll, 1978.

The USA and UK were the leading countries in the introduction of steamships. Steam propulsion accounted for one percent of total gross tonnage in the merchant fleet of these countries by 1817* (USA) and 1825 (UK). With the exception of France, where further detailed analysis of the causes for the significantly longer diffusion time would

---

* Particularly in the early (turbulent) phase of the introduction of a new technology, some deviations of empirical data from model estimates can be observed. Based on Table 3.1.1 the *estimated* introduction dates (1 percent market share) are 1827 (+2 years) for the UK and 1810 (-7 years) for the USA.

Table 3.1.1. Share of steam/motor ships in gross tonnage of commercial ship fleet 1818–1938.

Country	t_0	Δt years	Estimated Market Share 90%	99%
Austria +	1890	47.6	1913	1937
France	1898	62.6	1929	1960
Germany *	1894	41.9	1915	1936
Russia °	1899	55.8	1927	1955
UK	1884	57.0	1913	1941
USA	1885	75.0	1923	1960
World	1894	55.2	1921	1949

+ Austria–Hungary 1829–1913
* German Reich 1850–1913
° Tzarist Russia 1859–1913

[Source: UK (Nakicenovic, 1984); USA (Montroll, 1978); all estimates for other countries are based on data from Mitchell, 1980, and Woytinsky, 1927.]

be required, the USA and England as early starters have also the longest diffusion time constant Δt. However, their lead becomes progressively absorbed, and in the end most countries enter their saturation periods (after penetration beyond the 90 percent market share) at surprisingly close time intervals.

Thus, the time interval between the appearance date of steamships at a significant level (i.e., above 1 percent market share), spanning in our sample 35 years (1817 in the case of the USA and 1852 for Germany), becomes reduced to 16 years between the first (Austria–Hungary: 1913) and the last country (France: 1929) reaching their saturation phase (i.e., market shares exceeding 90 percent). This time lag gets somewhat widened to 24 years when considering the completion dates of the substitution process (i.e., sail ships accounting for less than 1 percent of gross commercial fleet tonnage). This, however, does not appear important for two reasons. First, substitution proceeds very slowly above the 90 percent market share level, (as we speak about the flat asymptotic last part of the S-shaped substitution trajectory) and offers only marginal market share gains. Second, as discussed above, steamships in turn started to be replaced by motor ships at the beginning of the 1920s (see Figure 3.1.5). Thus the 90 percent penetration dates can be considered as a fair estimate of the

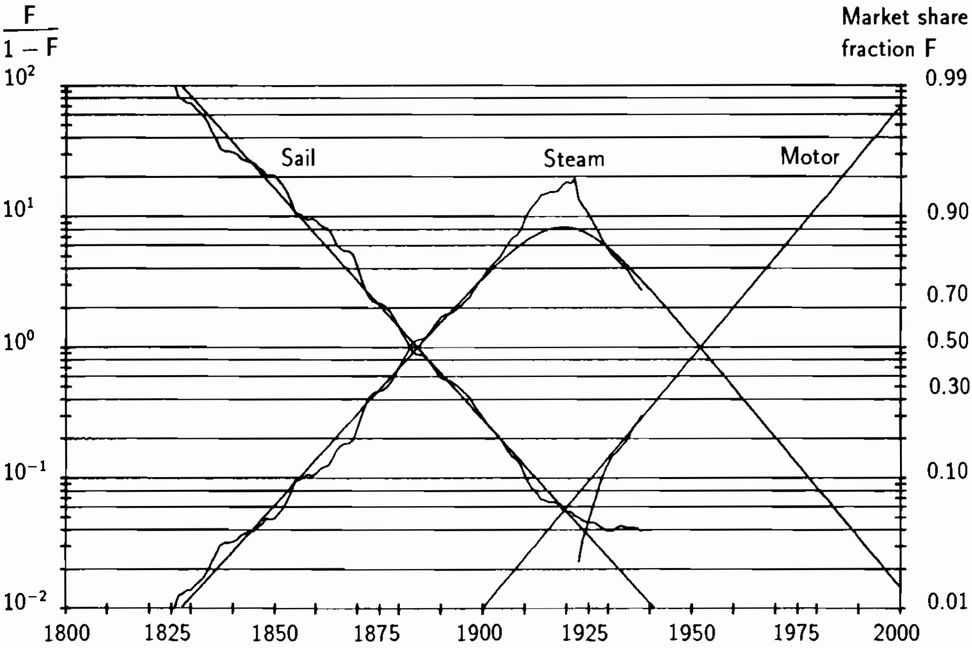


Figure 3.1.5. Share of sail-, steam-, and motor ships in gross tonnage registered in the commercial ship fleet of the United Kingdom. (Source: Nakicenovic, 1984.)

saturation of the effective replacement of sail ships by *steamships* proper.

It is our contention (and we will try in subsequent sections to substantiate this hypothesis by further empirical evidence in other growth and substitution "clusters") that we observe here a "catch-up" effect in the diffusion patterns between the countries who began early and countries following later. Thus, although, the dates of introduction (or start of diffusion/substitution) may be spaced over time, all countries (or technologies within a "cluster") reach their saturation phase within a relatively short time period, due to this catch-up effect. We would like to call this synchronization of the completion of diffusion and substitution processes the *season of saturations.*

The economic implications of this saturation synchronization, especially when it occurs for a number of technologies (products) in different countries, could be formulated in the following way. The rapid replacement of old technologies by new (e.g., sail ships by steamships), especially in the time period between 10 to 90 percent penetration (i.e., the steep part of the S-shaped diffusion/substitution trajectory) generates growth impulses due to the large investments into new equipment and through the induced growth of the industries manufacturing this equipment. In other words, economic theory argues (as discussed in Section 2.2.2 above) that diffusion of a new technique and the related replacement of old techniques, results in market expansion to a new (equilibrium) level.

Once this process reaches saturation (say beyond the 90 percent penetration level), expansion becomes almost stationary. Demand expansion (new steamship orders to replace sail ships) stops, leaving only replacement demand to be satisfied. In addition, the new level of market expansion (which became possible through the technical, economic, etc., possibilities opened up by a new technology) is asymptotically reached.

Even if a new technology is introduced at this stage (and in fact, the next technological "cluster" frequently develops during the saturation phase of the previous one), it will only have a small economic impact until its penetrates beyond the 10 percent market share level. Significant growth impulses on the economy can therefore only come from a new technology "cluster" after it enters its main growth phase. A "technological cluster" in the present context consists of a host of interrelated new technologies, products, organizational forms of production, and mediating socio-institutional innovations. Furthermore, the "cluster" is driven by developments in a number of core countries following (catching up) the leaders (early starters) in diffusion. These core countries form an international "diffusion bandwagon", as they are linked by technology, human resources, information, and capital flows. Consequently the core countries show a similar trajectory of technology development and a resulting similar dynamic in the long-term structural change of their economic base.

The synchronization in the diffusion of steamships in replacing sail ships, especially in the saturation phase, constitutes empirical evidence for such a "bandwagon" in the area of transport technologies. To this end let us summarize the dynamics and especially the synchronization of the saturation of the "canal age" and the "steamship age". Figure 3.1.6 presents an account of the growth/substitution pulses for canals and steamships. Whereas not all countries reach saturation within the same expansion cycle (note in particular the data on Russia and later the USSR), the apparent congruence of the saturation periods of infrastructure growth and technological substitution in the transport sector is striking and provides an incentive to proceed in our analysis, to see whether other technologies match the pattern sketched out in Figure 3.1.6. As we shall see, they *do*.

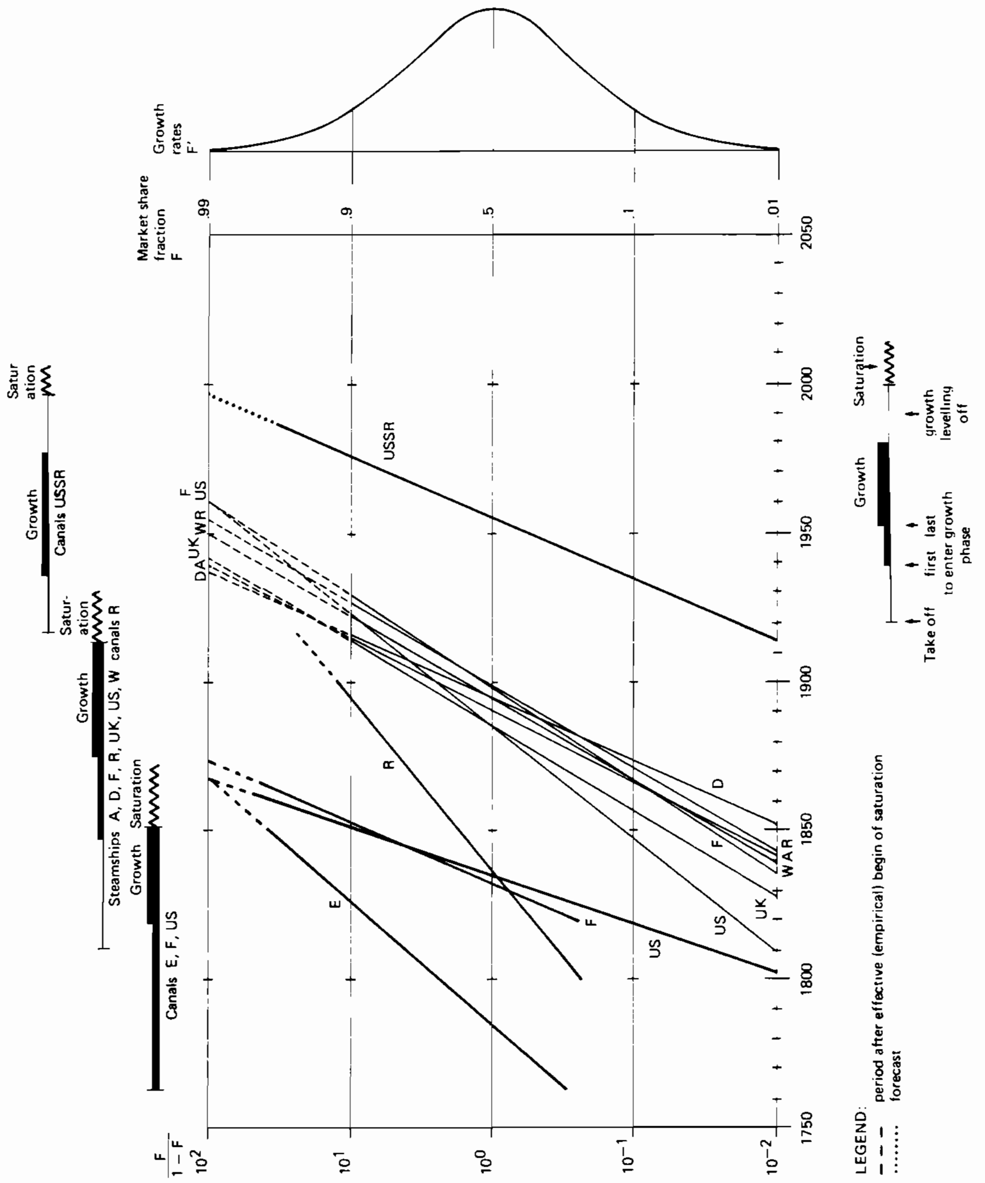


Figure 3.1.6. Growth pulses in the expansion of canal networks (c) and in the diffusion of steamships (s). *A* – Austria-Hungary (s), *D* – Germany (s), *E* – England (c), *F* – France (c and s), *R* – Russia (c and s), *UK* – UK (s), *US* – USA (c and s), *USSR* – USSR (c, second growth pulse), *W* – World (s).

3.2. Railways

The analysis of railway development is a field well covered by economic historians. The amount of narrative, be it pictorial/anecdotal, or more in-depth quantitative studies on the development and economic impact of railways, has grown to such volumes that a mere bibliography would fill considerably more space than the whole of this work. As a subjective choice of examples of excellent quantitative studies, we refer to the international compendium edited by O'Brien, 1983, and at the national level to the studies of Fogel, 1964, for the USA; Fremdling, 1975, for Germany; and Hawke, 1970, for the UK.

Consequently, we concentrate on a statistical account of the comparative *dynamics* of railway development in different countries, an area which, at least to our knowledge, has not received a great deal of attention (one exception being Mothes, 1950). In addition, we expand our holistic view of railway development in analyzing the whole *life cycle* of railways. Thus, we not only consider growth, but also the subsequent saturation and decline in various countries. Finally, examples of technological change in railways for different countries are presented when analyzing the replacement of steam by diesel/electric locomotives.

The long-term pattern of railway expansion

The growth of the railway network, following the opening of the 20 km Stockton & Darlington Railway in 1825* is analyzed like the other infrastructure systems in this study, viewing railway expansion as a diffusion process interacting with an (ultimately limiting) environment. The take-off of the railways demonstrates the importance of the *simultaneous* and *complementary* character of infrastructure and technology development. When these two streams of development

---

* We retain this as the starting date of railway expansion merely out of convenience. As Geise, 1959, rightly points out, the development of steam railways started considerably before George Stephenson. Railways are the result of developments in the English coal mining industry (see e.g., Nef, 1932): the wooden and later, iron-rail wagonway for overland coal transport as well as the steam engine for mine water drainage. For instance, at the opening of the first French railway line (17 km between St. Étienne and Andrézieux) in 1828, over 140 km of horse railways were in existence in the coal mining area of the Loire alone (Stürmer, 1872), a figure reached by the French national railway network only 6 years later. A good description of early (pre-1825) railway development in England is given by Marshall, 1938. In terms of importance for the transport system the beginning of the railway era in England should possibly be 1830, i.e., the opening of the Manchester-Liverpool railway line. Note also, that the length of the railway network in England in 1834 was only one percent more than the final (maximum) network size (32,800 km in 1928). Stürmer, 1872, and Voigt, 1965, provide a good overview of early railway development in countries other than the UK.

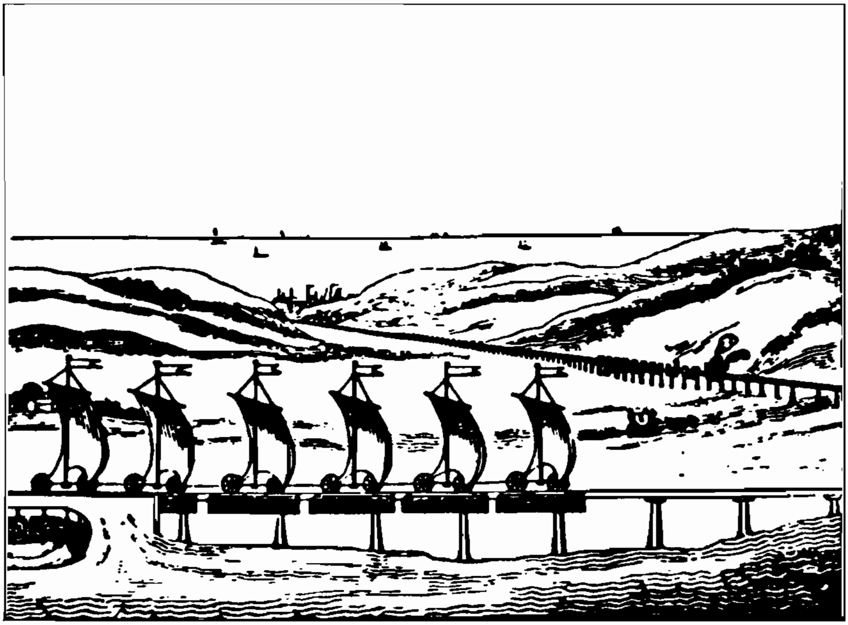


Figure 3.2.1. An innovation that failed in absence of a new technological (prime mover) base. Monorail railway using sails, as proposed by Henry R. Palmer in 1828. (Source: Marshall, 1938.)

merged together (i.e., the combination of the iron wagonway with the steam engine as the prime mover), a technological quantum leap opened, resulting in the take-off of a new infrastructure system. Thus, the synergistic whole is more than the sum of the parts. If either one of the two developments (infrastructure or technology base) takes place in isolation, significant macro-economic effects of a new technology appear unlikely. The history of technology provides many examples illustrating this point, as shown in Figure 3.2.1. In the same vein, a quantum leap in the quality of road transport became only possible after the introduction of the internal combustion engine as the prime mover. Before that date, even ambitious road construction programs were not able to significantly improve the slow transport speeds of horse-drawn carriages and wagons. Could this provide an historical analogy for judging the possible success and effects of the currently proposed ambitious railway infrastructure reconstruction programs in a number of European countries? Only the future will tell, whether this kind of scepticism is appropriate or not. Let us return instead to

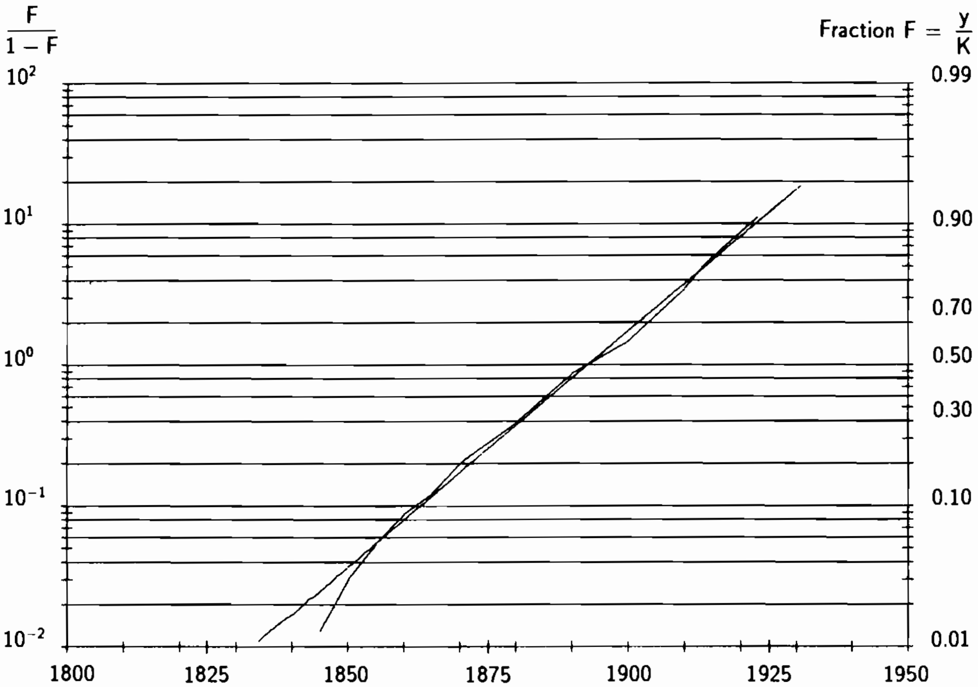


Figure 3.2.2. Growth of the world railway network [K 1.323 10^6 km, t_0 1893, Δt 57 years]. (Data source: Stürmer, 1872, and Woytinsky, 1927.)

the beginning of the railway era and consider the expansion of the railways in a number of countries. The diffusion of railways can be analyzed along two dimensions: time and space.

An analysis of the diffusion process of railways in *time* is reported in Figure 3.2.2 for the expansion of the world railway network. The diffusion process appears – with the exception of some minor deviation prior to 1855 – to evolve regularly along a logistic trajectory. Total railway network length increased with a Δt of 57 years to over 1.3 million km worldwide, saturating at the end of the 1930s. The somewhat faster growth at the beginning of the diffusion process can be observed in many countries. However, in Figure 3.2.2, such a deviation appears somewhat insignificant* occurring only in the early turbulent phase of the diffusion process (below 5 percent of final K). We note, however, cases in which such an asymmetry appears indeed significant, in particular in the case of the UK, although it is mostly due to the fact that official railway statistics ignore private mine railways, which was an

---

* Sensitivity analysis performed with alternative S-shaped diffusion models (Gompertz curve and generalized Sharif-Kabir models) did not improve significantly the (already high) R^2 value (0.998) of the logistic fit and shows also higher deviations of the estimated saturation level from the empirical data.

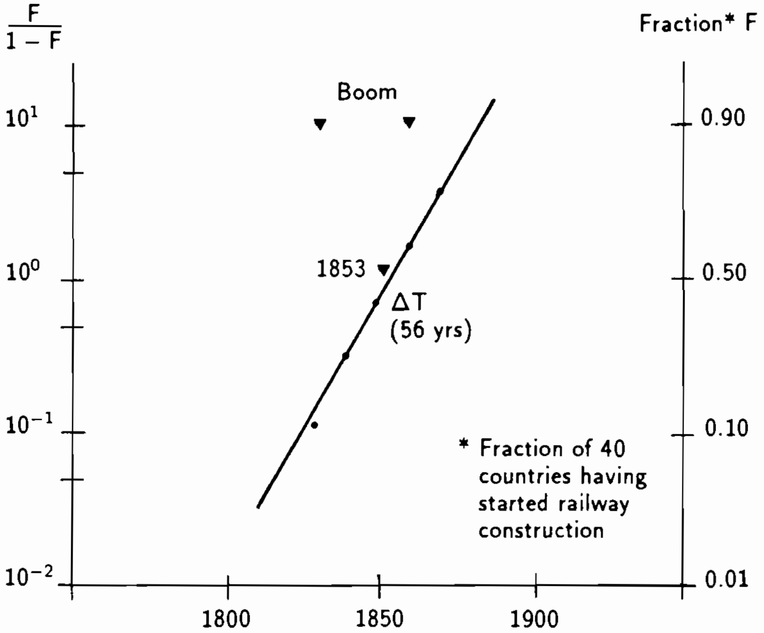


Figure 3.2.3. Diffusion of starting dates of national railway network construction. The time lag between the starting date of railway construction in follow-on countries up to 1825 is considered as a *functional* measure of their distance to the innovation center (England). Note that by 1900 virtually all national railway networks in the world had begun. (Source: Marchetti, 1987, see also Table 3.2.1.)

important (precursor) infrastructure in the early railway era. Appropriate asymmetric growth models will be tested to describe the development pattern of railways for such cases (as in Table 3.2.2 below).

The second dimension of the expansion of railways is their diffusion in *space*. Here, diffusion can be first analyzed as the spread of an infrastructure system from the innovation center (England) to peripheral regions (other countries), by recording the construction start-up dates of national railway networks, as shown in Figure 3.2.3. In comparison to Figure 3.2.2, we see that the dynamics of the temporal and spatial diffusion process of the railway network are identical (Δt of 57 and 56 years respectively). However, the start-up pulse precedes the growth pulse proper by some 40 years. A corollary of the identical dynamics in the spatial (as represented by the start-up pulse) and the temporal diffusion processes is, that the growth of the railway networks of individual countries proceeded via an international

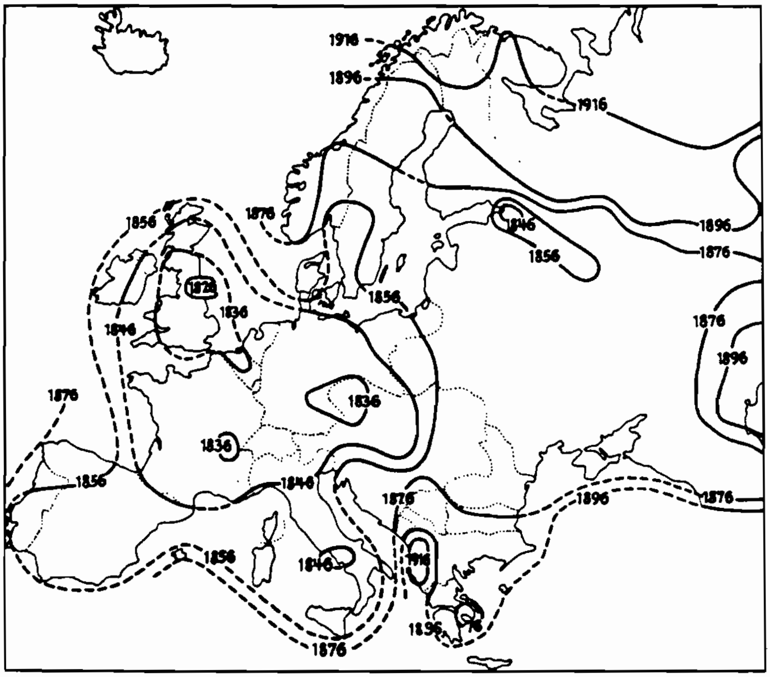


Figure 3.2.4. Spatial diffusion of the European railway network, in isolines of areas having railway networks by a given date (in 10-year time intervals). (Source: Godlund, 1952.)

"bandwagon" effect. This resulted in the near simultaneous saturation of railway expansion by the 1930s in all industrialized countries.

This conclusion is confirmed when we look at the *spatial diffusion* of railways proper, i.e., by analyzing the spread of the railway network in terms of spatial coverage and through a hierarchy of innovation centers as shown in Figure 3.2.4. Here the regularities of the hierarchy and neighborhood effect in the spatial diffusion of innovations (Section 2.3) can be observed. The innovation spread out from the innovation center (1826) to its hinterland (all of England and parts of Belgium) as well as to sub-innovation centers (St. Etienne-Lyon in France, and Austria-Bohemia* in central Europe) by 1836. From this hierarchy of innovation centers the innovation "wave" had spread out further by 1846: from England to Scotland, Wales, and part of Ireland and from the continental innovation centers (Austria and France) to most of central Europe, as well as to a third level in the hierarchy of innovation centers (Russia, in 1838 with the opening of

---

* The first railway line on the Continent, according to Stürmer, 1872, was the horse railway from Linz to Budweis constructed between 1825 and 1832.

the St. Petersburg–Zarskoje Selo line). From 1846 on, the innovation spread further to the European hinterlands, and by 1876 all of central Europe, the largest part of European Russia, southern Scandinavia and part of the Balkans were covered by railway networks. The spatial innovation wave finally reached its last (4th) innovation center level: Greece. From that date onwards, the wave started to ebb and progressed to the most remote hinterland areas from the 3rd and 4th hierarchy sub-innovation centers.

The completion of the expansion of railway networks in Europe happened in peripheral areas (e.g., in Finland) to some extent later than in the core countries, but occurred no later than the 1940s. This implies that the growth process in peripheral areas, where railway construction began at a later date (up to 50 years later than England), proceeded much faster. Therefore, the analysis of the spatial diffusion of railways as illustrated in Figure 3.2.4, confirms a significant catching-up tendency for latecomers. As we show below, these latestarters do not, however, develop their infrastructure system in any way as extensively as those countries that started early.

Table 3.2.1 summarizes the growth of the railway network in a number of countries (discussed individually in more detail below), along with their estimated diffusion model parameters and the resulting *densities* per unit of country area and per capita. With the exception of Tzarist Russia (catching up with a Δt of 37.4 years) the time constant of diffusion (i.e., to go from 10 to 90 percent of ultimate saturation length) of the railway network ranges from 47 to 57 years for the countries in our sample. The build up of railway infrastructures was thus a rather long process, but of comparable duration in the countries analyzed.

With the exception of the second phase in railway construction in the USSR, railway network expansion was completed in all countries in Table 3.2.1 by the 1930s. The fact that the length of the railway network in Austria-Hungary and Germany peaked in 1913 does not contradict such a statement because we have to take into account the territorial changes (losses) of these two countries following World War I. Bearing this in mind, we can conclude that the railway expansion in these countries also saturated around 1930. The closeness of the saturation dates in railway expansion is all the more noteworthy considering their long growth period of nearly 100 years. A quick look at the available statistics (Mitchell, 1980) suggests that a similar statement can be made for an even larger sample of European countries. The railway network reached its maximum length around the 1930s: including the Netherlands (1929), Denmark (1932), Belgium (1933), Switzerland (1937), Sweden and Greece (1938), and finally Italy (1940). A railway "diffusion bandwagon" thus existed for a large

Table 3.2.1. The growth of railway networks (km).

Country	Estimated Saturation Length (K)	Maximum Length Achieved	t_0	Δt	1% of Maximum Length	Maximum Length	Density[2] Length in 1925	
	1000 km	1000 km	year	years	year	year[1]	per 100 km²	per 10,000 Inhabitants
Austria-Hungary[3]	24.5	23.0*	1883	51.8	1841	1913*	8.0	10.2
France	42.5	42.6	1876	47.1	1841	1933	9.7	13.7
Germany[4]	66.8	63.4*	1882	57.0	1841	1913*	12.2	9.6
Russia	73.9	70.2*	1890	37.4	1851	1913*	0.3–1.5	4.8–8.4
USSR	147.8[5]	145.6	1949	43.6	1921	1986	0.7[5]	5.5[5]
UK[6]	33.9	32.8	1858	56.6	1834	1928	16.0	8.8
USA	526.1	482.7	1891	54.5	1840	1929	4.3	38.1
Core Countries in 1925[7]		715.0					2.0	16.7
Rest of the World in 1925[7]		540.0					0.6	3.7
WORLD	1322.9	1255.0[8]	1893	56.9	1844	1930	1.0	6.7

* Important territorial changes thereafter.

[1] Source: Mitchell, 1980.

[2] Density as calculated by Woytinsky, 1927, for 1925, except Russia and the USSR (own calculation). Range of figures for Russia corresponds to total density and the European part of the territory respectively. Density figures for USSR are for the 1986 network size.

[3] Austria alone after 1919.

[4] Including Elsaß-Lothringen (E–L); excluding E–L the maximum network size was approximately 60 10^3 km by 1938.

[5] Including intercept (73.9 10^3 km); 90% probability that K lies within 145–150 10^3 km. Density figures for 1986 network.

[6] Parameters refer to a Gompertz function ($t_0 = K/e$ instead of $K/2$).

[7] Core countries including Russia. Rest of the world is the difference between world total and core countries.

[8] Source: Mothes, 1950.

number of industrialized core countries, as presented in Table 3.2.1 (summarized in Figure 3.2.15).

A number of, at that time, still developing countries (e.g., Japan, India, Latin America and Africa, i.e., the late adopting countries represented in Figure 3.2.3 above) continued to expand their railway network even after the 1930s. However, as noted above, the expansion of the global railway network saturated at the level of 1.3 million km after 1930, and has remained constant ever since. Railway construction in peripheral (developing) countries exactly equaled the contraction of networks in industrialized countries. By 1930 the industrialized core countries represented in Table 3.2.1 had constructed over 715,000 km of railway lines, or some 57 percent of the world total. In 1986, these countries still used slightly more than 500,000 km. Thus, their share in the world railway network has decreased to 40 percent. Railway constructions in the periphery countries of the "railway bandwagon" represent thus a relatively modest effort compared to the scale and density realized in the core countries.

Noteworthy are also the different railway densities even among the various core countries. This illustrates, that the diffusion of infrastructure networks and systems, whilst being comparable in terms of their dynamics (Δts), will result in different absolute as well as relative density levels. Different diffusion levels (densities) are the result of the diverse geographical, economic, and social environments (boundary conditions) prevailing in different countries. In addition, the time period involved for the construction of the network, i.e., the time lag in the network construction start-up dates appears to be of importance. Consider just the density levels presented in Table 3.2.1. The spatial railway densities range between 0.7 km (USSR in 1986) and 16.0 km (UK in 1923) of railway lines per km^2, i.e., by a factor of more than 20. Per capita railway densities range from 5.5 km (USSR in 1986) to 38.1 km (USA in 1923) per 10,000 inhabitants, i.e., a factor of approximately seven.

However, Table 3.2.1 shows also, that the differences in the density levels between the core countries and the peripheral areas are even more divergent than between the core countries proper. Countries not participating in a particular infrastructure expansion phase achieve density levels significantly below those in core countries. Even in 1986, railway densities in areas outside the core countries of the "railway bandwagon" are, with a spatial density of 0.9 km railways per 100 km^2 of the country, and a per capita density of 1.9 km per 10,000 inhabitants, significantly lower than in the core countries. Railway densities in core countries continue to be higher, with values of 1.4 and 7.1 for the spatial and per capita railway densities respectively, despite the fact that these countries already decommissioned parts of their networks due to the development of new alternatives. Present railway densities of peripheral countries are even more divergent when compared to the density levels achieved in the core countries at the time of their maximum network size some 50 years ago.

With respect to railway densities it is therefore interesting to extend our analysis to those countries that were not part of the "railway bandwagon" of Table 3.2.1. In analyzing the diffusion level (railway density* per km^2) resulting from the growth of these, to varying degrees, "railway latecomers", we observe an interesting fact: The diffusion levels appear to decrease the later a country started the construction of a railway network, as suggested by Figure 3.2.5. Railway densities are calculated for the year of maximum network size, or the latest year data were available in cases railway construction still continues.

---

* Because of the strong population dynamics of these countries, per capita densities show even larger disparities compared to industrialized countries and are also decreasing for most developing countries.

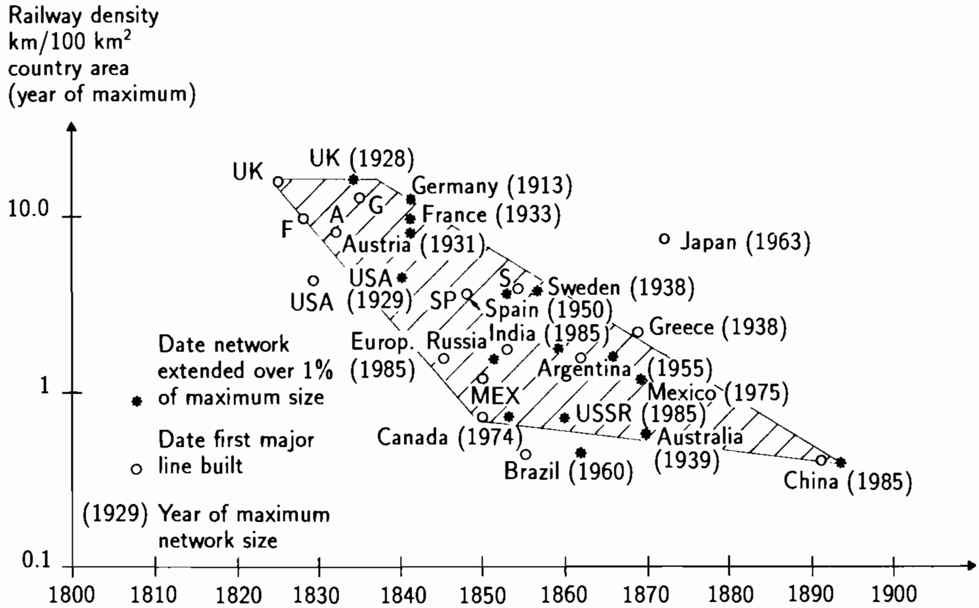


Figure 3.2.5. Spatial railway density envelope versus network construction date. (Data source: Mitchell, 1980; Stürmer, 1872, and Fischer Weltalmanach, 1987.)

Two empirical measures are developed to assess the starting date for the construction of a national railway network. First the year the network extended for the first time over one percent of its maximum size, calculated either from the historical dates indicated in Figure 3.2.5, or in cases such as China where the network is still growing, from the current size of the network. The second measure records the date when the first railway line of importance* was constructed. Figure 3.2.5 suggests, that the resulting railway densities can be regrouped by a declining "density envelope" as a function of the date railway construction began. The later a country started, the proportionately fewer railway lines (per unit of country area) were constructed. The only exception is Japan, where construction started rather late (in the 1870s), but where a network on a similar scale (density) to the industrial core countries forming the "railway bandwagon" was built. This is possibly an indication that Japan was developing as early as the second half of the 19th century, the infrastructural

---

* For Tzarist Russia we put this date at 1845, ignoring the 27 km from St. Petersburg to Zarskoje Selo opened in 1838 as being of minor importance to the national transport system.

endowments required for joining the "club" of industrial countries later.

Thus, any analysis of the diffusion of infrastructures has not only to account for the different boundary conditions prevailing in different countries as represented in their size and geography, but also for the timing and pace a country follows in the diffusion of a particular technological system. There appears to be a "season" or opportunity window for the large-scale development of infrastructures. Developments started outside such a "window" will result in significantly lower growth and diffusion levels than those achieved in the core countries originating and synchronously developing a particular technological cluster. One could also consider that the USSR (as shown by its different infrastructure development pattern) was to a large extent "decoupled" from the expansion of the new technological base (the "car age"), which emerged in the 1930s, and which became a principal driving force of the post World War II growth in industries like automobiles, petrochemicals, consumer durables, etc., in western market economies. China would have to increase its present railway network by more than a factor of 20 in order to achieve a similar spatial railway density that the industrialized core countries attained in the 1930s. This does not only appear infeasible for practical reasons, it also would be simply absurd to suggest repeating* a growth trajectory of a 19th century development phase.

Railway growth in individual countries

The expansion of the railway networks of the UK and the USA is discussed separately for two reasons. First, because of the importance of these two countries in railway development, either as the pioneer (UK) or as the country with the largest railway network in the world (USA). By 1840 US railways already extended over 4,500 km, while in Europe only some 3,000 km had been opened altogether (Taylor, 1962). By 1930, the saturation of railway expansion, the USA alone accounted for over 480,000 km of railways, compared to around 400,000 km in Europe and 1,255,000 km worldwide (Woytinsky, 1927, and Mothes, 1950). This dominance by the USA (close to 40 percent of the world's railways) is in fact not a unique historical situation. At

---

* This does not of course imply that developing countries should not construct any new railway lines. For the transport of low-value goods, say soy beans or iron ore, railways are an efficient and cost effective transport mode for these countries. Our conclusion just refers to the fact that we do not consider successful catching up of these countries to rely in any way on the development of an infrastructure which was so essential for the growth of industrial countries in the 19th century.

Table 3.2.2. Sensitivity analysis: growth of UK railways.

Algorithm	Estimated K	% Deviation of K from max. length*	R^2
UK (10^3 km) 1825–1912 (whole growth period):			
1	30.876	-6.00	.9914
2	32.341	-1.54	.9877
3	–	–	–
4	33.944	+1.51	.9932
5	–	–	–
6 $\gamma = 1$	37.393	+13.84	.9820
6 $\gamma = 0.7$	36.490	+11.11	.9759
6 $\gamma = 0.5$	–	–	–
6 $\gamma = 0.3$	–	–	–
6 $\gamma = 0.1$	–	–	–
UK (10^3 km) 1855–1912 (main growth period):			
1	32.924	+2.37	.9938
2	33.924	+3.28	.9923
3	34.544	+5.17	.9919
4	35.794	+8.98	.9951
5	–	–	–
6 $\gamma = 1$	41.667	+26.86	.9960
6 $\gamma = 0.7$	40.652	+23.68	.9957
6 $\gamma = 0.5$	39.735	+20.97	.9952
6 $\gamma = 0.3$	38.497	+17.20	.9944
6 $\gamma = 0.1$	–	–	–

* Maximum length in the UK: 32.85 thousand km (in 1928).

– no numerical convergence of algorithm

Algorithms: 1-3 logistic [1 non-linear least-squares fit (LSQ), 2 non-linear LSQ with data dependent weights, 3 linear LSQ], 4 Gompertz (linear LSQ), 5 modified exponential (linear LSQ), 6 Sharif-Kabir model with different "retarding factor" γ (linear LSQ).

present the USA accounts for close to 40 percent of the total cars registered worldwide (MVMA, 1983).

The second reason for considering the UK and the USA in more detail is methodological. As mentioned above, the expansion of the railway network proceeded somewhat faster in the early phase (pre-1850) than it would have if the diffusion process had adhered to a perfectly symmetric growth trajectory. In Chapter 2, we discussed the fact that some researchers claim that diffusion should indeed proceed along a positively skewed S-shaped pattern. As a result, a number of asymmetrical diffusion models (Gompertz, Sharif-Kabir, modified negative exponential) have been proposed in the literature. The case of the UK and the USA therefore provide an empirical test of various asymmetrical diffusion models and this is summarized in Tables 3.2.2 and 3.2.3. Railway diffusion is analyzed retrospectively for two periods, for the whole diffusion period, and for the main growth period. The different algorithms discussed in Section 2.1 are tested and two measures of goodness of fit are presented: The deviation of the estimated saturation level K from the maximum network size in the two countries (USA: 300 10^3 miles by 1929; UK 32.85 10^3 km by 1928), and second, R^2 values as a statistical measure* of goodness of fit.

United Kingdom (UK)

The growth of the UK railway network (excluding Ireland), when the whole data period is considered, is apparently best described by a Gompertz function, both in terms of the model deviation from the empirical final saturation level and in terms of the R^2 of the estimate. This confirms a certain asymmetry of the growth pattern of the UK railway network in its early phase. The logistic non-linear least square algorithm with data dependent weights also performs well in terms of the two goodness of fit criteria, however, it overestimates the very early growth (pre-1845) phase of the UK railway expansion (see Figure 3.2.6). The performance of other asymmetric diffusion models, in particular the "flexible" generalized Sharif-Kabir model turned out to be rather disappointing, resulting in a considerable overestimation of the saturation level. For the modified exponential model, no numerical convergence of the algorithm was achieved (i.e., no model fit to the empirical data was possible). When considering data for the main growth period alone (i.e., after the turbulent early growth phase), the fit of the logistic function appears to describe the development pattern more precisely (especially in terms of the estimated saturation level) than the Gompertz model. Again the performance of the Sharif-Kabir model turned out to be inferior to the logistic and Gompertz models, despite marginally higher R^2 values and the modified exponential model failed again to reproduce the empirical data base.

---

* Note here however, that R^2 values differ only marginally and are comparable only to a limited extent, due to the fact that for algorithms 1) and 2) residuals between the actual and estimated data are calculated, whereas for algorithms 3) to 6) residuals of (linear) *transforms* of data from the estimate are measured. Note that R^2 values refer only to data segments used for parameter estimation, as indicated in Tables 3.2.2. and 3.2.3.

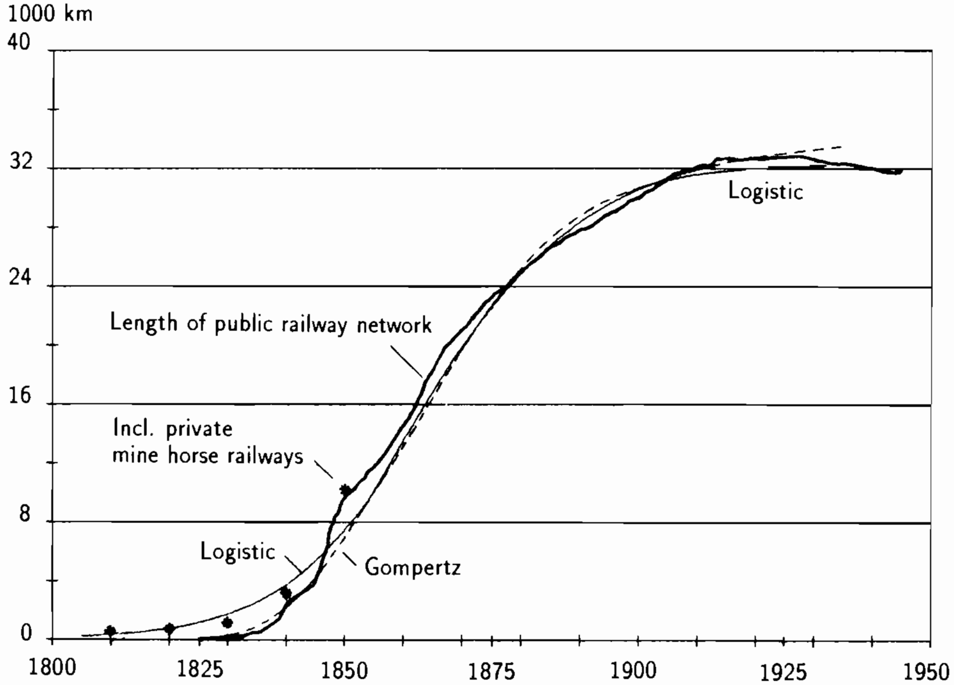


Figure 3.2.6. Growth of the railway network of the UK (in 1000 km) and diffusion models based on logistic and Gompertz curves (for parameters see Tables 3.2.1 and 3.2.2). Data points marked separately refer to total railway network length including private mine horse railways. (Data source: Mitchell, 1980, and Pollard, 1988.)

One can thus conclude for the UK that the Gompertz model is preferable for the *ex post* description of the historical process, but does not perform as well* as the logistic model as a forecasting tool (i.e., estimating the final saturation level K with incomplete data). Particularly noteworthy is the relatively poor performance of the generalized Sharif-Kabir model, and the inappropriateness of the modified exponential model, even in a case depicting relatively strong asymmetry of the diffusion process. This led us to enquire into the possible reasons for such a strong initial expansion of the network. As it turned out, a good part of the asymmetrical diffusion pattern is due to a definitional problem of the quantitative measure of the diffusion process, as the official statistics only recorded *public* railway lines. Especially in the early phase of railway development, and in particular before the advent of Stephenson's steam locomotive, a large number of private mine wagon and rail-ways existed, using wagons drawn by

---

* See in particular the larger variation in the estimated K with shorter data segments in Table 3.2.2.

horses for coal transport. If one includes these early systems into the figures (as in Figure 3.2.6 using estimates by Pollard, 1988) we see that in the early phase of railway development the rail-based transport infrastructure was considerably larger, in which case a less skewed diffusion model appears more appropriate as a descriptive tool. This points to a more general measurement problem with which diffusion research is confronted. Usually, early precursor or "hybrid" systems are not included in the analysis, but it appears that precisely these transitional systems may hold the clue to explaining the rapid initial adoption phase and thus a skewed diffusion pattern. This is also the reason why the development of the new French rapid train system (TGV) will be considered separately in our analysis of the French railway infrastructure below.

United States of America (USA)

The conclusions with respect to the appropriate diffusion model used to describe the growth of the railway network of the USA (Table 3.2.3) are as follows: The process is apparently a symmetric diffusion process, most adequately described by a logistic model, whereas all other models (including the Gompertz) result in significant overestimation of the saturation level of the diffusion process. Thus, despite a slight deviation in the early phase, before settling into a regular diffusion pattern (an observation made already by Fisher and Pry, 1971, in their analysis of technological substitution processes), the logistic appears to be the most appropriate model to describe the growth of the railways, and has consequently been retained for the subsequent analysis of other countries.*

The growth of the railway network of the USA proceeded after the opening of the first line, the Baltimore-Ohio Railroad in 1829, with a "diffusion half-time" Δt of 55 years, and reached the maximum network size by 1929. The smooth evolution of the railway expansion along the logistic trajectory (see top of Figure 3.2.7) somehow masks the considerable engineering (see e.g., Taylor, 1962) and financing difficulties (see for example, Poor, 1860, Goodrich, 1960, and Myers, 1970) which had to be overcome, including strong social resistance, as illustrated in the bottom of Figure 3.2.7.

In view of these technical, economic, and social complexities, the regularity of the growth process in the USA as depicted in Figures 3.2.7 and 3.2.9 is indeed striking. Figure 3.2.8 gives a pictorial

---

* Sensitivity analysis for all these countries yielded similar results as in the case of the USA.

Table 3.2.3. Sensitivity analysis: growth of USA railways.

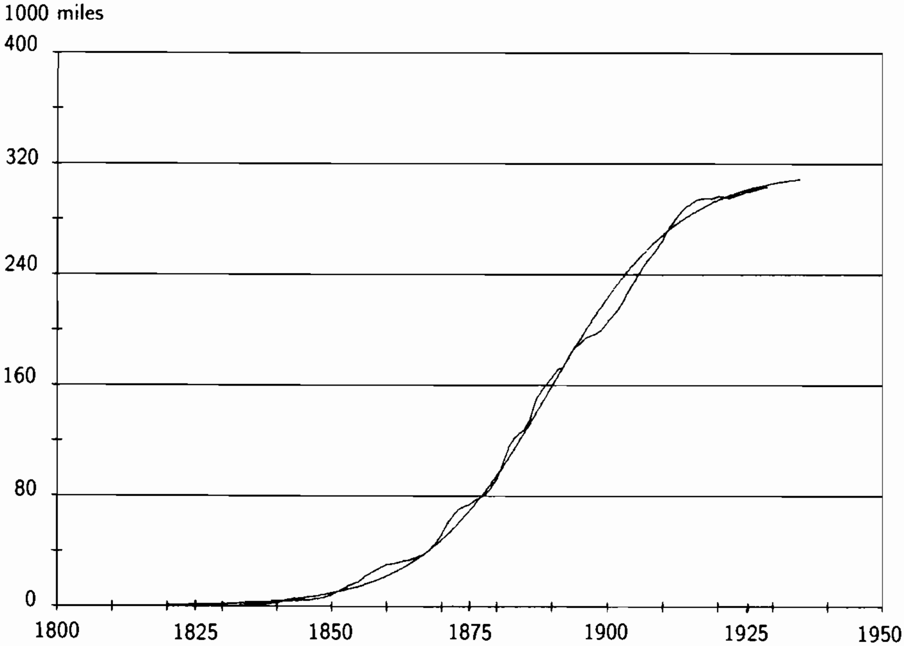
Algorithm	Estimated K	% Deviation of K from max. length*	R^2
USA (10^3 miles) 1830–1920 (whole growth period):			
1	332.983	+10.99	.99558
2	327.040	+9.01	.99668
3	308.543	+2.09	.96236
4	495.629	+65.21	.99588
5	–	–	–
6 $\gamma = 1$	406.168	+35.39	.97570
6 $\gamma = 0.7$	384.580	+28.19	.97323
6 $\gamma = 0.5$	368.581	+22.86	.97106
6 $\gamma = 0.3$	350.350	+16.78	.96824
6 $\gamma = 0.1$	327.222	+9.07	.96437
USA (10^3 miles) 1855–1920 (main growth period):			
1	332.983	+10.99	.99558
2	337.980	+11.27	.99544
3	341.557	+14.23	.99485
4	549.080	+83.03	.99240
5	–	–	–
6 $\gamma = 1$	492.545	+64.18	.99510
6 $\gamma = 0.7$	458.029	+52.68	.99516
6 $\gamma = 0.5$	433.185	+44.39	.99514
6 $\gamma = 0.3$	403.887	+34.63	.99500
6 $\gamma = 0.1$	367.960	+22.27	.99493

* Maximum length USA: 300 thousand miles (in 1929).

– no numerical convergence of algorithm

Algorithms: see Table 3.2.2.

illustration of the somewhat abstract numerical representation of the expansion of the US railways presented in Figures 3.2.7 and 3.2.9. The expansion of the railway network in the USA started to move from East to West and resulted finally in very different infrastructural densities between different parts of the country. The contrast between the high density of railway lines in the rich agricultural Midwest and the industrial Northeast (in particular the Great Lakes area and Chicago), and the low railway density in the West is particularly striking. Incidentally, a similar pattern can also be observed in



MOTHERS LOOK OUT FOR YOUR CHILDREN!

ARTISANS, MECHANICS, CITIZENS!

When you leave your family in health must you be hurried home to mourn a

DREADFUL CASUALITY!

PHILADELPHIANS your RIGHTS are being invaded! regardless of your interests or the LIVES OF YOUR LITTLE ONES. THE CAMDEN AND AMBOY, with the assistance of other companies, without a Charter and in VIOLATION OF LAW, as decreed by your Courts are laying a

LOCOMOTIVE RAIL ROAD!

Through your most Beautiful Streets to the RUIN of your TRADE, annihilation of your RIGHTS and regardless of your PROSPERITY and COMFORT. Will you permit this? or do you consent to be a

SUBURB OF NEW YORK!!

Rails are now being laid on BROAD STREET to CONNECT the TRENTON RAIL ROAD with the WILMINGTON and BALTIMORE ROAD, under the pretence of constructing a City Passenger Railway from the Navy Yard to Fairmount!!! This is done under the auspices of the CAMDEN AND AMBOY MONOPOLY!

RALLY PEOPLE in the Majesty of your Strength and forbid THIS

OUTRAGE!

Figure 3.2.7. Growth of the railway network in the USA 1830–1930 (top); (Source: Table 3.2.1 and 3.2.3). Resistance to the growth of railways in the USA (bottom). (Source: January 1838, Courtesy of Metro-North Commuter Railroad, New York.)

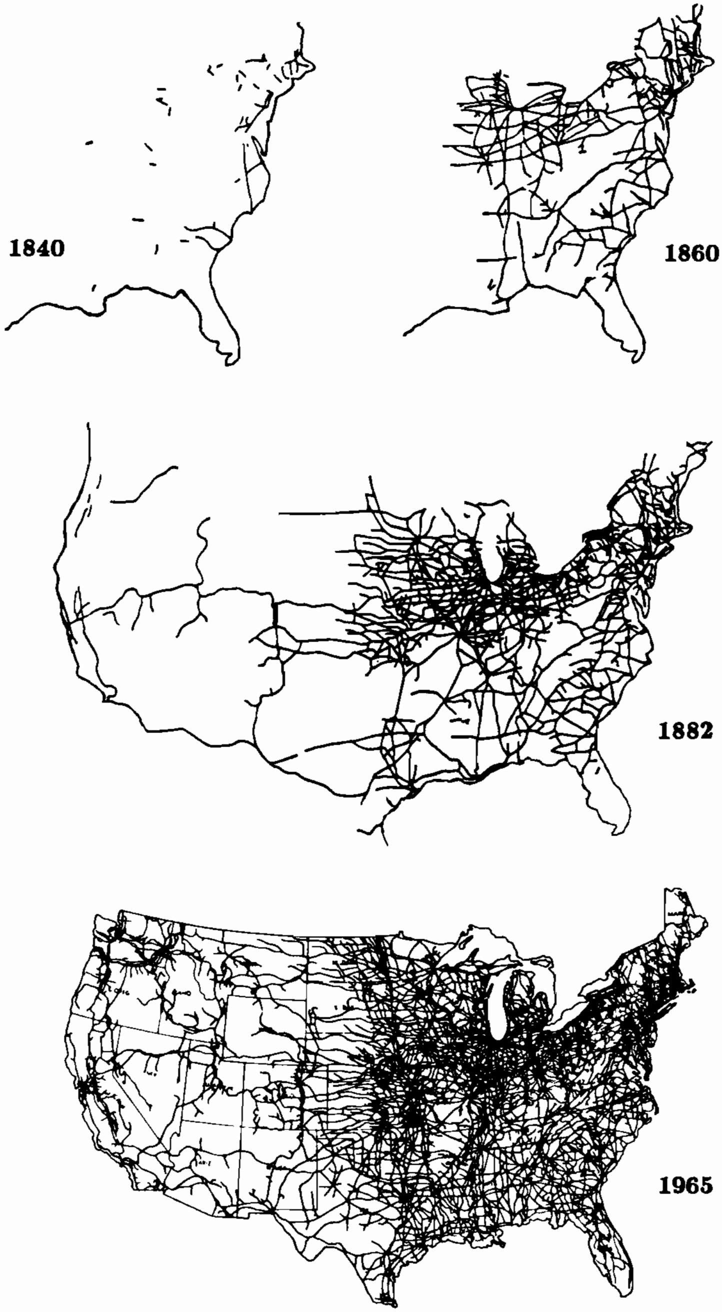


Figure 3.2.8. Railway network of the USA in 1840, 1860, 1882, and 1965. (Source: adapted from Lord and Lord, 1953, and Morrill, 1970.)

the USSR, with higher railway densities in the European USSR, compared to the Asian part of the country.

We have noted above that the railway network of the USA reached its maximum size in 1929. Since that date, the length of the network has decreased, representing the last phase of the technological life cycle of the railways: their *decline.*

The life cycle of railways

Up to now we have discussed railways only during their expansion phase. Below we extend the analysis up to the present in order to provide quantitative evidence of both the rise and the fall of railway networks (measured by their size) in industrialized countries. We do not comment here on the causes of saturation and the subsequent decline of the railway networks. These will be discussed in Chapter 4, where the development of railways is considered together with the evolution of other transport infrastructures and technologies.

USA, UK, and Germany (FRG)

Figure 3.2.9 presents the life cycle of railways in the UK, the USA, and in Germany. A strikingly similar pattern emerges: The railway network appears to grow along a regular logistic pattern (the asymmetric growth pattern of the UK is approximated by two logistic epochs) and to contract thereafter according to a logistic trajectory with possibly zero asymptote. As we are still in the early phases of the contraction process, and in any case prior to the inflection point (half of the maximum network size), the statistical uncertainty* of the estimates is so great, that the forecasts presented in Figure 3.2.9 are speculative in nature.

USA

The estimate of the railway life cycle for the USA (data: US DOC, 1975 and 1989) reveals a symmetry in the growth and decline of the US railway system. The contraction process proceeds, however, with

---

* The statistical uncertainty of the asymptote of the contraction process is (as the upper bound is known, i.e., the network cannot contract below zero level) primarily a function of what percentage this process has progressed up to the present. As all networks have still not yet reached or passed the inflection point of the contraction process (half of the maximum network size) the statistical uncertainty of the final asymptote is substantial. Based on the results of Monte-Carlo simulation of parameter uncertainty, presented in Chapter 2, we estimate that there is a 90 percent probability that the final railway network will be between 0 to 54,300 miles in the USA (179,000 miles main track operated in 1986); between 0 and 3,000 km in case of the UK (16,670 km in 1986); and between 0 to 18,200 km for the FRG (30,568 km in 1985).

a Δt of 116 years, two times slower than the growth process (Δt of 55 years). Still, the regularity of both processes is noteworthy, pointing to a similarity in the driving forces in the rise and fall of railways. Positive and negative diffusion therefore characterize the life cycle of railways in the USA as well as in other countries. The structural *discontinuity* marking the transition from positive to negative diffusion occurred prior to 1930. In terms of absolute railway mileage, the USA has decommissioned the greatest amount of railway length (some 124,000 miles between 1929 and 1986). As a symbol for this process, the closure of regular transcontinental passenger train connections stands out, since the market for long-distance passenger transport was entirely lost to public air and private car transport. In terms of relative closure however, the USA (having closed down roughly one third of its railway network) ranks second after the UK, where the railway network shrunk to about half of its maximum size.

UK

The situation in the UK (data: Mitchell, 1980, and DOT, 1987; Ireland is excluded in Table 3.2.1 and Figure 3.2.9) appears to be similar to the USA, with two noteworthy exceptions. First, the growth process exhibits a certain asymmetry (at least for the public railway infrastructure). Second, the contraction process is, with a Δt of 51 years much faster than in the USA or the FRG. In addition, the process does not always evolve along a smooth regular pattern. The deviations around World War I and in the 1920s appear to be important according to Figure 3.2.9. This however, is due to the particular linear (logit) transformation employed here, expanding deviations near to the saturation level. More noteworthy is, however, the "overshooting" of railway closures between 1960 and 1970, and the marginal decreases from 1970 to the mid-1980s. Therefore in order to estimate the parameters of the contraction process we have only used data from the period 1928 to 1960. As it turned out, the 1960 model forecast proved quite accurate for the length of the railway network in 1986.

Germany (FRG)

A life cycle analysis of the railway network length of Germany and later the FRG (data: Mitchell, 1980, and Stat. B.A., 1987) turns out to be rather complex. Whereas the growth process depicted in Figure 3.2.9 proceeds along a regular logistic path as in other countries, one has, however, to take into account the territorial losses of the German Reich after World War I and the state and geographical discontinuity

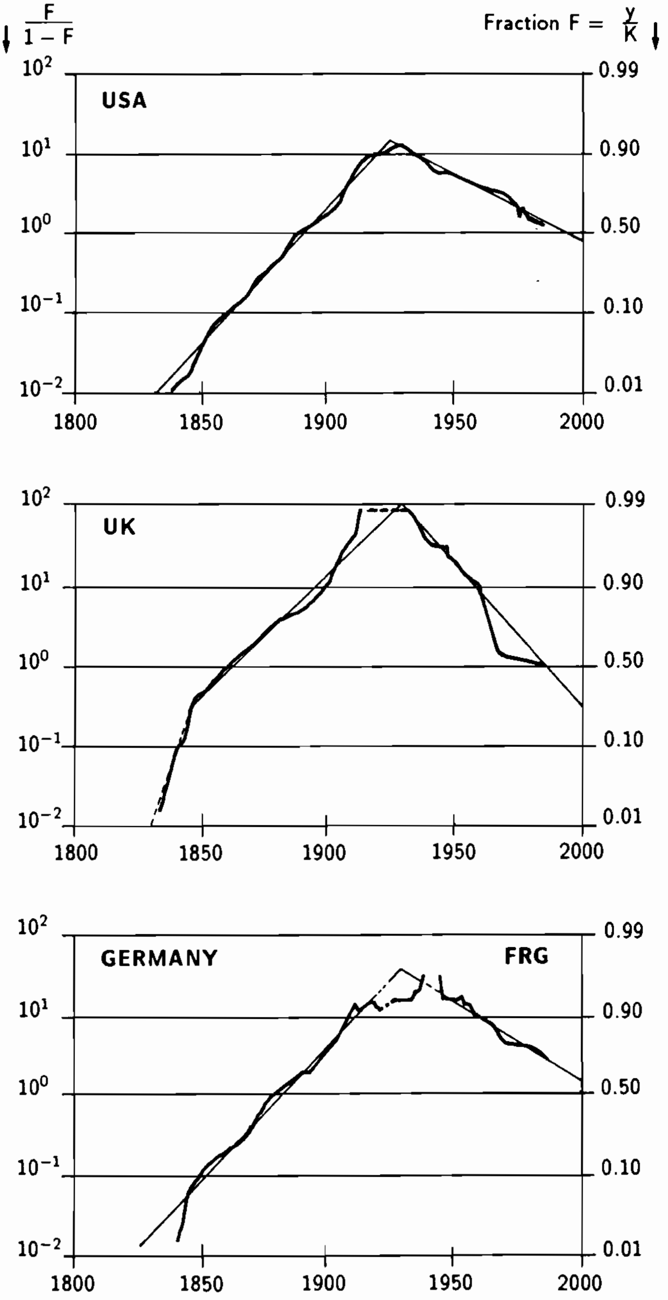


Figure 3.2.9. Life cycle of railway network length in the USA, UK, and Germany (FRG). For growth parameters see Table 3.2.1, estimated decline parameters [USA: t_0 1996, Δt 116 years; UK: t_0 1986, Δt 51 years; FRG: t_0 2011, Δt 106 years].

after World War II. If we make a rough estimate of the share* of the railway network within the new boundaries as compared to the old ones, we see a similar picture emerging as for the USA and the UK. Again functional symmetry between positive and negative diffusion (with the latter having a considerably longer time constant, Δt of 106 years) prevails. The discontinuity date for the transition from positive to negative diffusion occurred by 1930, as suggested in Figure 3.2.9. Thus, the process of decline was initiated at a much earlier date, and is *not* the result of post World War II transport policy, as frequently argued.

Austria, France, and Russia/USSR

Our analysis of the life cycle of railways would not be complete if we did not present counterexamples to the picture emerging from Figure 3.2.9. This is shown in Figure 3.2.10, where the railway life cycles for Austria, France, and Russia/USSR are presented. All these countries are of special interest, because of the different policies pursued with respect to railway development after the saturation of railway expansion in the 1930s. In Austria, the length of the railway network has remained practically constant since 1919 (around 6,000 km). France experienced a major societal and technological discontinuity (discussed in more detail in Chapter 4) as a result of being occupied during World War II, and is currently pursuing an ambitious railway construction program for the high speed TGV train. Finally, the USSR has experienced two major discontinuities (the October Revolution of 1917, as well as the devastating impacts of World War II) and is the only industrial country to pursue an extensive railway construction program, doubling the length of the railway network compared to Tzarist times.

Austria

The constancy of the railway network of Austria (data: Mitchell, 1980, and ÖStZA, 1987) provides a good counterexample to the USA, the UK, and Germany (FRG). Since the 1930s, only marginal closures (after taking into account the territorial changes and loss of railway lines of the Monarchy) have been effected. The contraction tendency as suggested in Figure 3.2.10 is thus entirely speculative as this process has progressed only to the level of the first few percent, and might

---

* K, as a basis for Figure 3.2.9 is renormalized (i.e., is changing over time) to account for the change in country (and railway network) size.

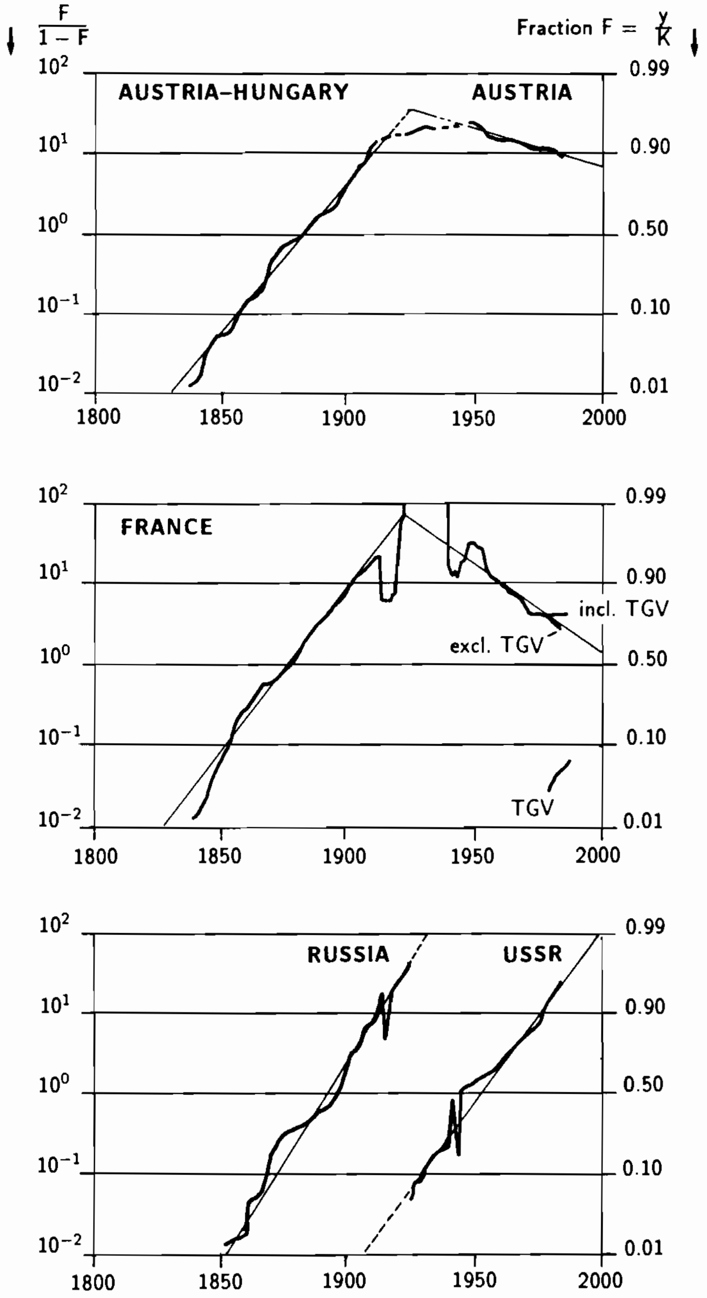


Figure 3.2.10. Life cycle of railway network length in Austria, France, and Russia/USSR. For growth parameters see Table 3.2.1, decline trends are speculative.

not be statistically significant* at all. Probably the most striking result of our speculation depicted in Figure 3.2.10 is the estimated Δt of 195 years for the process of decline. At least in the case of Austria, railway lines appear to be here to stay. The cause of this has frequently been attributed to the strong representation and influence of the railway trade union in the political system of Austria. Of course, the economic performance for a large part of the network is a different story. In a recent public debate on the possibility of closing some regional railway lines it was noted that the transport costs (to the Austrian nationalized railways) per passenger-km on certain lines exceed the high fares of Viennese taxis.

France

The case of the French railway life cycle is particularly interesting, because of the large perturbations of usually smooth diffusion trajectories caused by external events, i.e., the two World Wars. The growth in the French railway network (Mitchell, 1980, and Ann. Stat., 1987) proceeded along a regular (if we disregard the effects of World War I) diffusion path, which after 1850, is matched quite accurately by the logistic pattern indicated in Figure 3.2.10. This process is, with a Δt of 47 years, somewhat faster than in other countries, where 55 years is typical. Maximum network size was attained in 1933 with 42,600 km (100 km above the estimated saturation level K of the logistic diffusion process) and remained at that level up to 1939 (therefore data are outside the ordinates of Figure 3.2.10). During World War II the railway network shrank by about seven percent and the resulting reconstruction phase lasted until around 1950 when the network extended again to 41,300 km. Since that time the railway network has decreased to the 1986 level of 34,640 km, or by 18.7 percent. Using the time period of 1955 to 1986, this decline process is estimated** to proceed with a Δt of around 98 years. Without going into detail about the plausibility of extrapolating this process into the future, we point to an interesting discontinuity. The decline process, extending over 30 years, appears to have nearly stopped or, at least to have been reduced to a very small rate since 1980. Total network size remained nearly constant at a level of around 34,600 km. This

---

* We estimate that there is a 90 percent probability that the final railway network will be between 0 and 7,800 km. The higher range figure exceeds even the length of the present day system. Hence, we cannot conclude a statistically significant contraction tendency in the data on the length of the railway in Austria.

** The TGV lines put into service after 1981 were excluded in the parameter estimation data, they are plotted separately in Figure 3.2.10. Based on the results of parameter uncertainty of the model, there is a 90 percent probability that the final network size will be between 0 and 21,000 km (1986: 34,640 km). Thus, the statistical uncertainty involved in such an estimate is substantial.

discontinuity coincides with the appearance of the TGV high speed train network.

The emergence of the TGV may be interpreted from yet another angle: France is the European country with the highest share of nuclear electricity in it's energy balance. As such the TGV may represent a first attempt to substitute oil with nuclear electricity in the transport sector. Although the network served by the TGV is still relatively small in comparison to the total railway network size (Figure 3.2.10), it is expected to grow further (TGV Atlantique and further connections to central Europe). An interesting development to be monitored within the next decade will be whether the new railway policy, as manifest in the TGV, will be successful in halting the decay process of the traditional network, or whether the TGV network will expand at the expense of traditional railway links, concentrating railway traffic on just a few high-capacity connections.

The expansion and the success of the TGV connections in France, which partly uses its own lines, and partly uses the existing, traditional railway network, has been frequently seen as an indicator of a "railway renaissance" in Europe. However – like many early pioneering systems – the TGV combines elements of both traditional and of new (forthcoming) infrastructures. As such, it is functionally probably similar to early railway lines (like the *horse* railway between Linz and Budweis in 1832), and does not represent *the* ultimate model of a high speed, long-distance transit system for the next millennium. Finally, one has to keep in mind, that despite its high visibility and undeniable success* on certain routes (e.g., Paris-Lyon), its market impact is still relatively minor. At present, the TGV transports slightly more than one percent of the total passenger-km in France and is therefore still in the early "embryonic" phase of its life cycle development. Therefore future options with respect to the technological design of rail based, high speed, and throughput systems and especially with respect to the appropriate institutional/organizational frameworks for financing, construction, and operation of such systems, may still be open, as alternatives developed in the FRG or in Japan suggest.

---

* The success of the Paris-Lyon TGV connection can be best rationalized by the fact that the TGV altered the space-time activity framework of the people who commuted between the two cities, i.e., that the TGV became the *fastest* transport mode between Paris and Lyon. Its future growth will thus depend on whether this can be repeated on other connections. As the recent (December 1989) speed record on the Atlantique line indicates, this can indeed be achieved for new connections within France. However, in view of the institutional rigidities of European railways, it is at present unclear if similar time budget economies, especially in comparison to air connections, can be realized on a European scale. Interface problems (offsetting the large part of the speed gains of the TGV network proper) with the UK and with the FRG still remain unresolved. Again this is not a technological question but rather one of institutions.

USSR

As our final counterexample to the decline of railway networks in industrialized countries, we consider railway development in the USSR (data: Mitchell, 1980, and Narod. Khoz., 1982, and 1986). Following the expansion of the railway system of Tzarist Russia, saturation appears to be in tune with the life cycle of the railways of other European countries. The newly created USSR initiated, however, an ambitious second railway development program in the 1930s, although many of the projects planned at that date were only started much later. The most prominent example is the second Trans-Siberian railway line Baikal-Amur Magistral (BAM).

Figure 3.2.10 shows that in the second phase the railway network of the USSR doubled in size from around 71,000 km in 1918 to 145,600 km in 1986. This increase followed a logistic trajectory with a Δt of 45 years, and was thus similar to the first growth pulse. The second growth pulse proceeds, however, not along a regular path, but instead is characterized by a period of high volatility due to the effects of World War II. After the invasion of the USSR, there was a "burst" of railway lines. The total network increased from 86,400 km in 1939 to 106,100 km in 1940, i.e., by nearly 20,000 km. This was the result of an industrial relocation program to, and behind the Urals, and was made possible by dismantling a large part of the yet uncompleted railway tracks in Siberia and reusing them for railway construction in the western part of the country (Glaziev, 1988). This burst became, however, elastically absorbed in due course, and from the mid-1950s growth again seems to follow the logistic trajectory depicted in Figure 3.2.10. Railway expansion in the USSR appears to be currently approaching saturation.*

Let us conclude the discussion of the life cycle of railway development in different countries by describing the overall picture that has emerged from the quantitative analysis. We have noted that at the global level the length of the railway network has not increased since the 1930s. Hidden behind this global aggregation is in fact a great diversity in the railway network development pattern in different countries. In most of the core countries of the 1830 to 1930 expansion period, the length of the railway network declined; this tendency is most advanced in the UK and the USA. In other countries this tendency has been only marginal (Austria) or was brought to a halt, at

---

* As with all diffusion processes not yet completed, statistical measures of parameter uncertainty are given: assuming a 5 percent data error, the estimated saturation level K (146,600 km) of the diffusion process lies with 90 percent probability within 144,200 to 149,000 km (1986: 145,600 km). Parameters of the estimated diffusion model (see Table 3.2.1) are calculated based on data of the railway length in the period 1925–1939, and 1956–1986.

least temporarily, by a new railway policy (France). Only the USSR expanded the railway network after 1930. Since this expansion process in the USSR appears close to saturation, one would not expect the railway network of industrialized countries to remain at its present level, assuming that the contraction tendencies in most countries continue their historical trends. Also large-scale growth of railway infrastructures in developing countries to any levels close to those realized in industrialized countries appears unlikely. After about 50 years of constant network size at the global level, the traditional railway infrastructure of the globe is eventually facing a process of decline, even though the time constants involved, are very long, leaving room to consider new alternative technological, organizational, and institutional designs.

Effects of railways

Let us conclude this chapter on railways with its somewhat pessimistic future outlook and return to the "railway era" of the 19th century, in order to highlight the effects of the *growth* of the railways. This field is comprehensively covered by the so-called "new economic history" school and we will draw a few illustrative examples from this line of research. The economic effects of the railways have, indeed, been substantial despite the fact that railways do not qualify as the principal and single driving force of 19th century economic expansion. Railways were so successful that by the end of the 19th century they accounted for close to 90 percent of all passenger-km travelled and about 70 percent of all ton-km transported (as was the case in France; Toutain, 1967). Thus, the dominance of the railways in the transport system of the 19th century was significantly higher than the present dominance of road transport. Finally the competitive behavior and resulting pattern in the evolution of market shares of individual transport systems during this period (see Chapter 4) yields interesting insights into the dynamics of this process. These lead to a better understanding of what happened when, in turn, railways were challenged by new competitors, including the automobile.

The development of railways enabled (by their denser geographical coverage, higher transport speeds, and the lowering of transport costs) further specialization in production between industries as well as between geographical areas, which can be considered a major impetus for economic growth in the 19th century. Railways provided a novel, speedy, and after overcoming initial technological shortcomings, comfortable and safe means of passenger transport. O'Brien, 1983, notes for the 19th century that demand for travel seems to have been both price and income inelastic. Consequently, right from the beginning, railways captured (despite high initial tariffs) the highest value market niches in the transport sector: passengers and information (mail). For instance, passenger and mail boat service were soon discontinued after the appearance of railways in France. Even the Dutch, who had enjoyed with their *treckvaarten*, the most efficient mode of passenger transport by water, switched to rail, as soon as this facility became available.

The multitude of economic effects can be grouped into three broad areas. First, "forward linkages" in terms of influences on industry and regional specialization and the resulting impacts on urbanization and migration flows. In addition, decreasing transport costs changed prices and price differentials for basic commodities and finished products. Secondly, railways made direct demands on the industries supplying the products for their expansion in particular the construction, coal,* steel, and equipment manufacturing industries. These multiple effects can be summarized as economic "backward linkages". Finally, another impact of the railways was their effect on road and water transport and the accompanying drastic loss of market shares and sharp decline in profitability of the latter.

The "backward" and "forward" linkages of railways to the other sectors of the economy have been especially studied under the "leading sector" hypothesis (Rostow, 1952). The general conclusion of a number of empirical investigations (see e.g. O'Brien, 1983, Fishlow, 1965, Fogel, 1964, Fremdling, 1975, and Hawke, 1970) is that the railways were never the principal and only "growth locomotive" (or "panacea", O'Brien, 1983) for economic development in the 19th century. For the industry as a whole, railways absorbed – even at peak levels of construction (and induced** demand) – too low a percentage of total output to qualify as a leading sector *strictu sensu*. Railways are thus a *representative* element of the steam age rather than being *the* dominant (in terms of contribution to economic growth) technology cluster of the 19th century.

A number of historians have attempted to calculate the importance of the railways to national economies within the framework of neo-classical economic theory. This has resulted in estimates on the "social savings" of an economy due to the existence of railways. Despite all the theoretical and empirical imperfections (application of static equilibrium analysis to an essentially dynamic process of

---

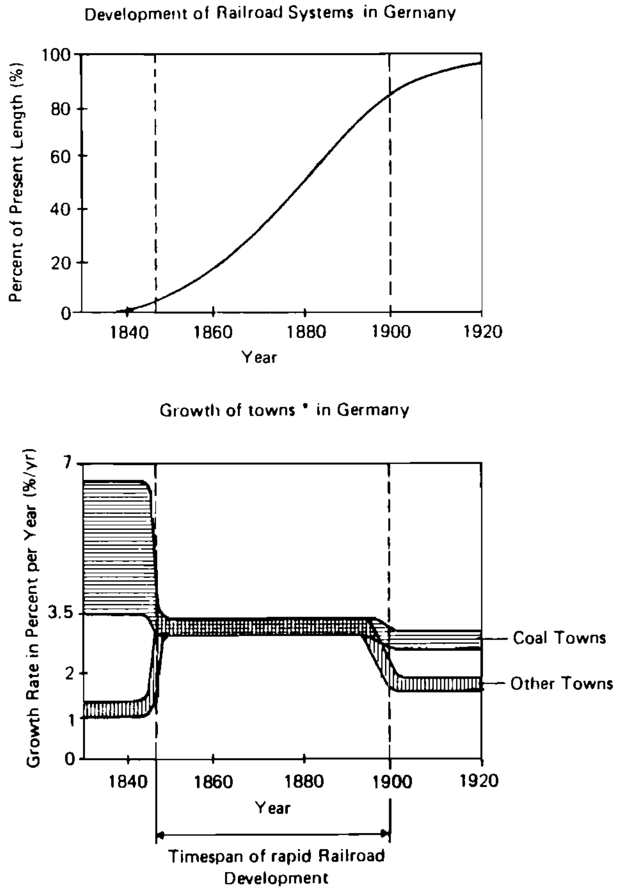
* Although the steam locomotive is usually associated with the usage of coal, one has to note that well into the 1870s the main fuel for steam locomotives in the USA was wood. Also the steam locomotives of Tzarist Russia were fueled to a considerable degree not by coal but with "mazout" (i.e., fuel oil).

** There were, however, important feedbacks to individual sectors e.g., the steel industry.

change) of this approach, it provides an interesting attempt to quantify the economic importance of the railways. The "social savings" associated with the existence of the railways is calculated by estimating the economic impacts (increase in transport costs) on a national economy, should the national railway network close down for one year. This *Gedankenexperiment* relies however on a stringent (and thus unrealistic) *ceteris paribus* assumption: that producers continue to dispatch exactly the same volume of goods and that passengers perform the same trips to the same destinations, even after the closure of the railway network. The "social savings" is thus the cost savings of an economy, compared to coping without the railways, expressed as a percentage of Gross National Product (GNP).

The conclusion of a number of studies calculating the "social savings" of the railways is that they varied between time (as a consequence of the increasing share of railways in transport volume) and space (as a consequence of different alternative infrastructure endowments). O'Brien, 1983, presents an overview of such estimates of "social savings" and concludes that they ranged typically between 5 to 10 percent of GNP for industrialized countries. However, the importance of railways in the industrialization process and changes in geographical structure transgresses this type of economic calculus. As an example of an important "forward linkage" of railways, their impact on settlement structure is presented in Figure 3.2.11. The unique growth rate of 3.5 percent/year of German towns between 1850 and 1900 was made possible by the development of the German railway network (main growth phase: 1853 to 1910, see Figure 3.2.9). The higher population growth rates of "coal towns" before 1850 and after 1910 point to the importance of coal, fueling the economic expansion pulse associated with the steam age in 19th century Germany.

Other important "forward linkages" of railways are that they widened markets (e.g. new export markets, Metzer, 1974), promoted economies of scale, helped (through regularity of service) to reduce inventories, encouraged the relocation of economic activities, stimulated competition, and increased trade and specialization. These effects were of special importance in countries not endowed with natural waterways. By lowering transport costs, railways changed the relative price structure and comparative locational advantage, especially of the primary sectors of agriculture and mining (including coal). Further significant market expansion of railways was enabled by progressively lowering transport tariffs as shown (in real terms) in Figure 3.2.12 for Canada. For France (Toutain, 1967, and Caron, 1983), Germany (Fremdling, 1975) the UK (e.g., Hawke, 1970, and Pollins, 1971) and the USA (Fogel, 1964, and Chapter 4) similar drastic reductions are typical: 25 percent (Germany), 30 percent



Development of Railroad Systems in Germany

Growth of towns * in Germany

* All towns with more than 50,000 inh. in 1850

Figure 3.2.11. Growth rates of cities in Germany during the phase of rapid railway development. (Source: Sassin, 1981.)

(Belgium) and 50 percent (France) lower tariffs are reported between 1845 and 1913. It is interesting to note, that this reduction proceeded slower in high-value market niches, where railways enjoyed a (technological and infrastructural) quasi monopoly situation, i.e., passenger transport.

For freight transport, the appearance of railways and the later reduction of freight rates intensified competition, forcing canal and water transport companies to reduce their tariffs accordingly, with a resulting decrease in their respective profitability as illustrated in Figure 3.2.13 for the case of the UK. Railways had two decisive advantages over inland navigation and canals in particular. First, the use of steam power enabled significantly higher transport speeds. Although steam power was soon introduced in inland navigation, canals never

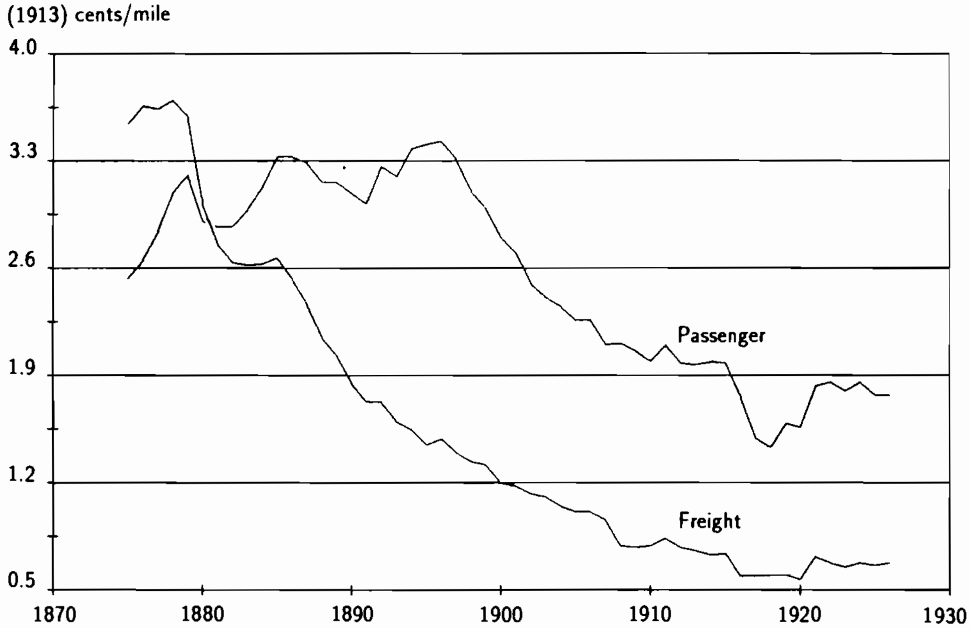


Figure 3.2.12. Real-term passenger and freight transport rates of Canadian railways in 1913 (Canadian cents per ton and passenger-mile). (Source: Green, 1986.)

could compete with the decisive speed advantage of railways. Secondly, the railway network developed very rapidly and provided for a significantly denser spatial coverage than was possible with the existing waterway and canal network.

Consequently, canal traffic immediately started to lose the highest value market niches: passenger and information (mail) transport* to the railways. Gradually, railways also conquered the mass commodity market (especially for coal**), despite fierce competition and tariff reductions by canal companies. As a result, the profitability of canal companies declined sharply and many canals were either closed down (as to a significant extent in the USA) or taken over by railway companies (as in the case of England).

---

* Quantitative evidence is presented in Part 4 in the analysis of the long-term evolution of the modal split of passenger traffic in France.

** Before the advent of the railways, the supply of coal to London was ensured by water (sea) transport. By 1845, railways transported 0.2 percent of the coal to London. Their market share increased within 30 years to over 65 percent (Mitchell and Deane, 1971).

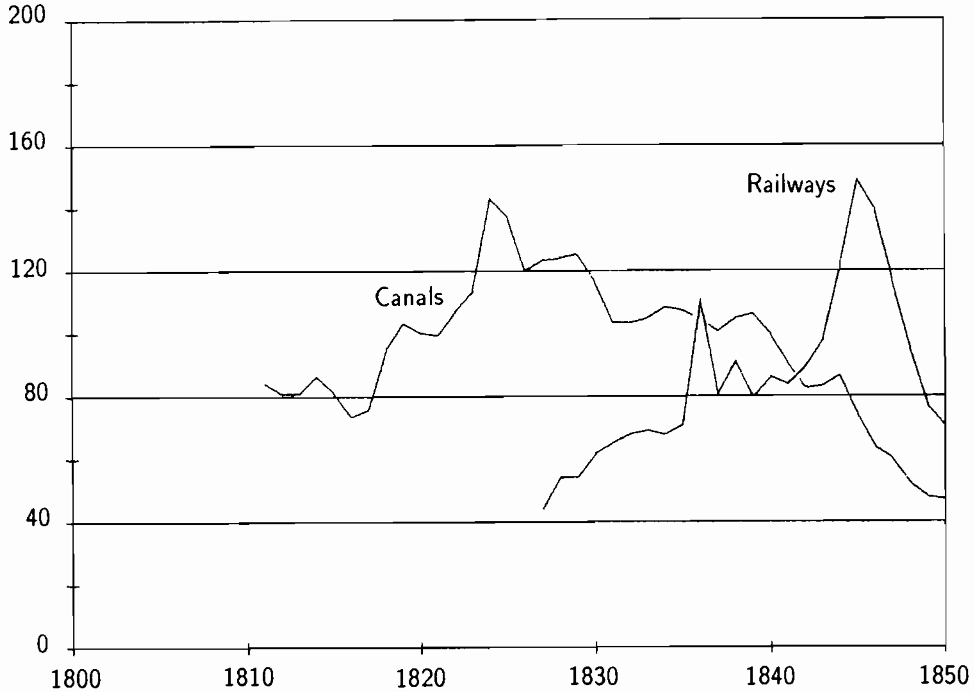


Figure 3.2.13. Yearly average index of canal and railway shares in the UK; June 1840 = 100. (Source: Enzelberger, 1989, based on Gayer *et al.*, 1953.)

The case of England is particularly interesting from an institutional perspective in terms of what happens to the competitor losing out on the market. Generally, the result of fierce competition between canals and railways (see Jackman, 1962) was a significant reduction in transport tariffs, which for England is estimated to range from a third to a half of the former canal transport rates. Dyos and Aldcroft, 1969, and Sax, 1920, provide more detail on the takeover of canal companies by railway companies in England. By 1870 already one third of canal mileage had been amalgamated with, or were controlled by, railway companies. A symbol for this was the takeover of the Bridgewater canal by a railway company in 1872. By 1883 already more than half of the canal mileage was controlled by railway companies. The result of railway and canal amalgamation was either the complete closure of canals and replacement by railways (frequently laid in the former canal bed) or in a market split between canals and railways leading to complaints about undue profits of "leviathan monopolists" (Dyos and Aldcroft, 1969). A number of government regulations in the form of parliamentary acts (1858, 1863, 1873, and 1888) were attempts to control the displacement of canals by railways, but in retrospect these efforts proved utterly unsuccessful. Could this provide a lesson for today's regulatory efforts to halt the further decline of the railways in many industrialized countries?

The decline of canals as a result of the success of railways is illustrated even more dramatically in the case of the USA. Only a few, high capacity, strategically located (e.g., where ocean ships could reach inland destinations) canals, like the Erie canal, survived the phase of extensive canal closures, reducing the length of the US canal network from some 7,000 km prior to 1870 to less than 2,900 km by the end of the 19th century.

Thus, water transport was progressively competed out of high-value market niches (passenger and information, i.e., mail, transport) and then gradually also from other market niches (first manufactured goods, then iron and coal), finally ending up with the lowest value market niche only (bulk commodities), where quality of transport (speed and regularity of service) does not play such a decisive role and fierce tariff competition prevails.

The competitive effects of railways on road transport were even more dramatic. The low speed and high costs (e.g., in France costs of road transport between 1850 and 1913 were on average three to four times higher than that of railways) of road transport meant that transportation by horses, mules, and oxen ended up by only providing complementary (distributive) services to and from the railroads and practically disappeared as an overland transport mode for passengers and goods. The already high cost differentials between road and rail transport at the beginning of the railway era widened even more over time, due to productivity increases (cost reductions) in rail transport, which were not matched by road transportation (Laffut, 1983). Currently, railroads now in turn appear to be trapped in this widening productivity gap, as their productivity increases have constantly been falling short of the productivity increases in the economy in general and of road transport in particular.

Although road transport improved during the 19th century, technical progress was rather incremental before the internal combustion engine provided a substitute for animal power used for cartage. Thus, the failure of road transport to compete with railways in the 19th century is a further corroboration of our hypothesis of the necessary synergistic evolution of infrastructure *and* prime mover technology. In the absence of a new (prime mover) technological base even significant infrastructural investments and improvements could not change the competitive position of a transport carrier in the long run.

To summarize, the impacts of railways on road transport in the 19th century can only be described as devastating. This can best be illustrated by the effects of railways on turnpike roads in England. By

1830 some 3,800 turnpike* trusts were operating around 32,000 km of (toll) roads in England and Wales. Their number was reduced to 1,048 by 1864; to 20 by 1886, and to 1 (!) in 1896 (Voigt, 1965). Consequently, road maintenance and new construction deteriorated, resulting in a further comparative disadvantage for road transport.

The competitive pattern outlined above stresses the importance of comparative speed and service advantages as well as of price insensitive, high-value market niches (passengers and information, later high-value goods) for the growth of a new transport system. These high-value markets were also lost first by railways to new competitors. By 1856 some 34 percent of freight receipts in England and Wales were obtained from low-value goods like minerals and lifestock (Hawke, 1970). Their share increased to around 50 percent by 1910 (Munby, 1978) and low-value goods accounted for between 65 to 71 percent of freight income in the period 1970 to 1986 (DOT, 1977, and 1987).

Thus the life cycle of railroads is characterized by penetrating first into the highest value market niches, and then conquering (via tariff reductions) progressively lower-value market niches. At the end of this process railways take the character of an indispensable (price and government controlled or owned) basic "commodity", becoming in turn vulnerable to competition from new transport systems, with improved quality characteristics. In perfect symmetry to their own success story, railways started to lose high-value market niches first (passengers to cars and mail to aircraft), and in due course were forced to concentrate more and more on lower value goods to be transported (as documented in Chapter 4).

We conclude with a comment on the dating and causality of railway saturation in the 1930s. We consider the saturation phenomenon of railways *not* to be a direct consequence of competition (from cars). As we will show below (Chapter 3.3), the automobile primarily grew by replacing horses as a means of road transport until the end of the 1930s. Cars were more *complementary* to railways, rather than being a competitor for long-distance travel. The saturation of the railways by 1930 should under such assumptions be rather interpreted as the approaching of a "feasibility limit", an exhaustion of the expansion potential of a "technological paradigm" named the railway or steam age, which in turn provided an opportunity window for the emergence of the automobile as a long-distance transport mode thereafter.

---

* Named after the turnpikes used for collecting the road tolls. This, however, faced considerable public opposition, resulting in frequent riots as for instance in Scotland and Wales (e.g., in 1843), as illustrated in Gordon, 1988.

Technological change in railways

The close association of railways with the technology of the steam engine is illustrated by an important technological change in railways: the replacement of steam by diesel and electric locomotives. This process is interesting for two reasons. First, it demonstrates that the decline in importance (both in terms of its infrastructure network and in terms of market shares) of the railways after 1930 cannot be attributed to a lack of technological innovation.

Second, the replacement process for steam locomotives is interesting from a methodological viewpoint as it provides evidence of a diffusion process with "negative" asymmetry. The initial penetration of diesel/electric locomotives was very slow up to the 10 percent market share, but then the process starts to converge to a regular logistic trajectory. As we have noted above in the development of the railway network, just the opposite effect (a "catch-up" effect to a logistic trajectory) can be observed. In any case, this phenomenon is worth noting as it provides a counterexample of an accelerating* diffusion rate, since it is often argued that diffusion should proceed with a decreasing rate (i.e., along one of the positively skewed diffusion models discussed in Chapter 2).

An interesting finding from Table 3.2.4 is the similarity in the dynamics of the replacement of steam locomotives among different countries. Δts range between 12.2 and 16.8 years (considering t-km as the most relevant indicator of substitution) and the inflection point of the substitution process lies (with the exception of the USA) within a close time interval somewhere between 1960 and 1964. The replacement of steam by diesel locomotives (electric locomotives have no practical importance in the USA) occurred some 10 years earlier in the USA than in other countries. This is illustrated when we compare the steam locomotive replacement process in the USA and the USSR (see Figure 3.2.14). The conformity of the replacement rates of steam locomotives in different countries, including a centrally planned economy, points to a similar comparative advantage for higher quality energy carriers (electricity either directly or indirectly via diesel) driving this substitution process.

---

* Note here that the last few percent of the substitution of steam locomotives proceeds faster than expected on the basis of a logistic model.

Table 3.2.4. Replacement of steam by diesel/electric locomotives.

Country	Measure for substitution	t_0	Δt	Estimated saturation 99% share	Empirical date of completion of replacement
Austria[1]	number	1967.2	15.1	1982	n.c.[2]
Austria[1]	t-km	1961.5	16.2	1979	1975
France	number	1961.9	16.6	1978	1972
FRG	number	1964.8	18.7	1983	1976
FRG[3]	t-km	1961.1	16.8	1978	1977
USA	t-km	1950.4	12.2	1963	1961
UK	number	1964.1	11.9	1976	1968
USSR	t-km	1960.4	12.6	1973	1973

[1] only Österreichische Bundesbahn (ÖBB)
[2] n.c. = not yet completed, market share of steam locomotives > 1%
[3] only Deutsche Bundesbahn (DB)

Data source: Austria: ÖStZA, 1975, 1980 & 1986; France: Ann. Stat. 1962 & 1975; FRG: Verkehr in Zahlen 1975 & 1985; USA: US DOC, 1975; UK: Munby, 1978; USSR: Kruglikov, 1985.

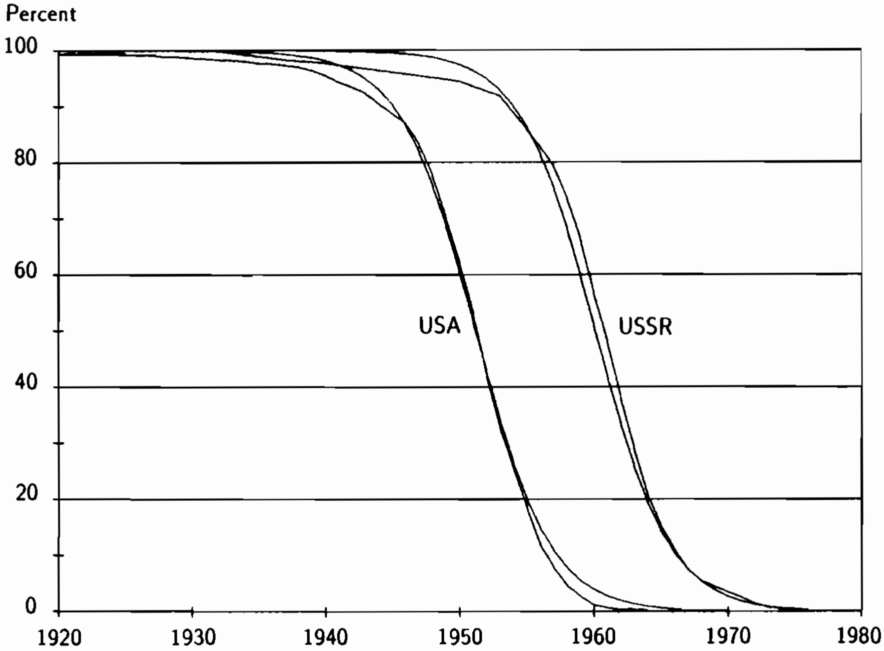


Figure 3.2.14. Replacement of steam locomotives in the USA and the USSR. (Source: Kruglikov, 1985 for the USSR; for the USA authors' estimates; for diffusion parameters see Table 3.2.4; data source: US DOC, 1975.)

Summary

Figure 3.2.15 presents an overview of the dynamics of the expansion of the railways and the replacement of steam locomotives. As with canals and steamships, a "bandwagon" phenomenon in the international diffusion of infrastructure systems and technological change in transport becomes visible. The expansion of railway networks in the industrial core countries is confined within a "diffusion frontier" defined by the diffusion processes of the first and the last country that developed a railway network. All diffusion curves of other countries lie within this diffusion frontier. The diffusion of railways in industrialized countries brought about a "quantum leap" in the productivity of the transport system, both in terms of speed and quality of transportation. Almost for the first time in history, growth in economic output and passenger travel was not constrained by an inadequate and expensive transport system.

The expansion potential of railway networks was realized up to the point where railways enjoyed a quasi technological and infrastructural monopoly position, transporting between 80 to 90 percent of all passenger- and ton-km in industrialized countries until it finally became exhausted. This in turn provided an opportunity window for the transition to new transport systems and infratsructures, in particular to road transport and the automobile. Our earlier hypothesis on the synchronization of the saturation periods appears to be further corroborated by Figure 3.2.15 on the basis of the diffusion of railways and later new forms of traction railways. A *season of saturations* can be seen in the period 1920 to 1930 and at the end of the 1960s and the beginning of the 1970s.

These dates constitute important turning points for the evolution of the transport system in all countries, marking the end of a period of expansion and market dominance, and the beginning of a period of decline both in terms of length of infrastructure grids and in terms of relative market shares. Such transition periods are frequently characterized by very volatile markets, as manifest in the performance of railway shares on the New York Stock exchange prior and after the 1929 crash, especially since railway shares were favored by speculators (Myers, 1970).

We will try to show in the following Section, that the technological cluster associated with oil usage and the diffusion of cars which emerged during the *season of saturations* of railways is, after about 50 years, approaching a similar phase of transition.

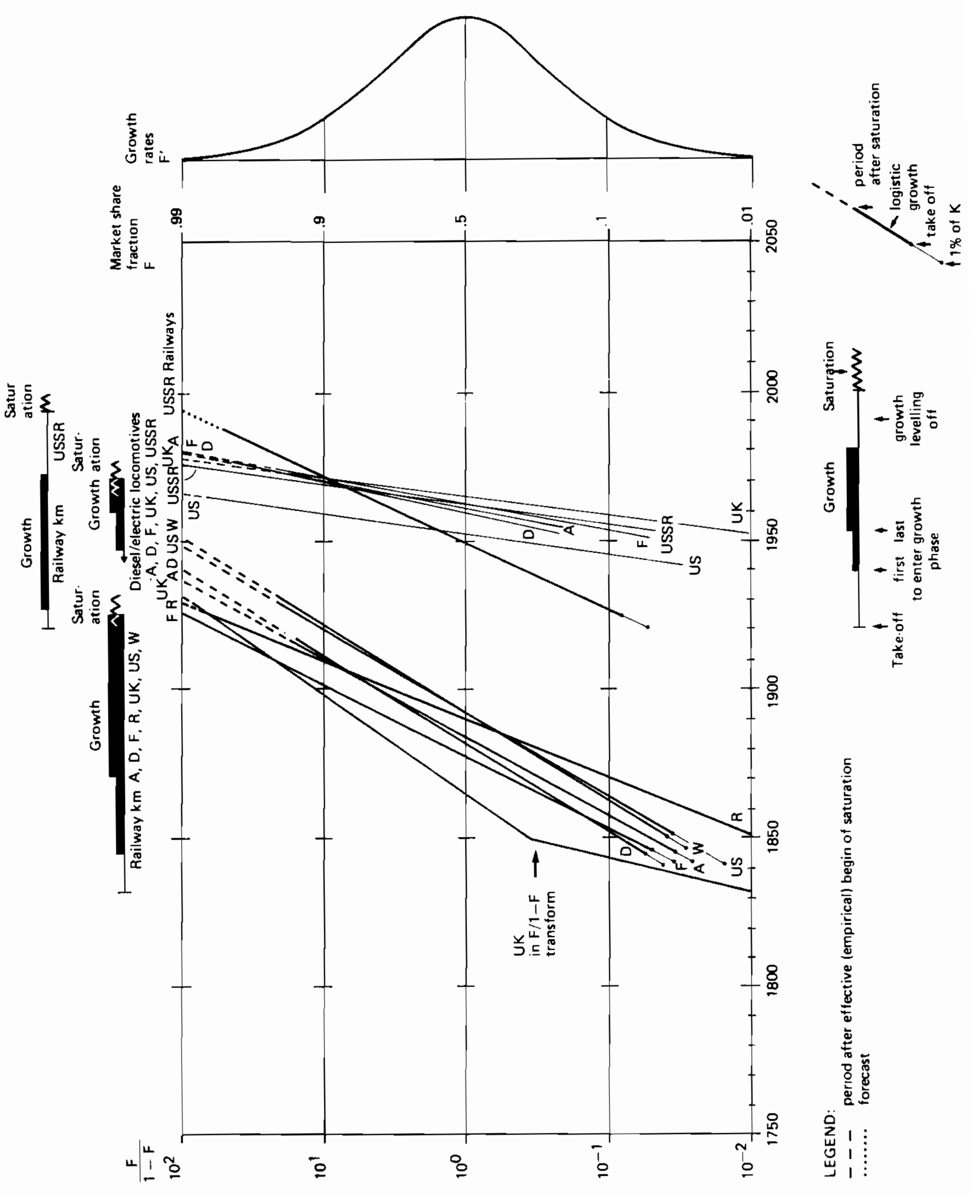


Figure 3.2.15. Growth pulses in the expansion of the railway network and in the replacement of steam locomotives. *A* – Austria-Hungary (later Austria), *D* – Germany (later FRG), *F* – France, *R* – Russia (first railway pulse), *UK* – UK, excluding Ireland, *US* – USA, *USSR* – USSR (second railway pulse, locomotive substitution), *W* – World.

3.3. Roads, Horses, and Cars

In this Section we discuss road transport under two major working hypotheses. The first concerns the relationship between road infrastructure and the appearance of the automobile. We argue, that the automobile did not create its own dedicated infrastructure as was the case with railways, but first made use of an already existing road network, which had been developed and improved in quality for horse driven carriages. Our second working hypothesis is based on the definition of two distinct phases in the diffusion of the automobile. These two phases are marked by different growth rates in the expansion of the automobile with a structural discontinuity occurring in the 1930s. We argue, that the first phase of the growth of the automobile was characterized by its competition within a relatively narrow market segment, i.e., the automobile was first diffusing into the market niches previously held by horse drawn carriages for passenger and goods transport. Its comparative advantage was particularly high in this market niche, which explains the replacement of horses, a process which was complete before World War II.

A corollary to this hypothesis is that the automobile did not enter into an effective competitive relationship with railways for long-distance travel before the substitution process for horses was nearing completion. In the early phase of automobile diffusion, cars had an essentially complementary role to railways, improving the transport services to and from railways and providing local (rural and urban) distribution functions. Two developments mark the transition into the second growth phase of the automobile. Technological improvements in the car itself, in particular the wide diffusion of closed car bodies by the end of the 1930s (enabling comfortable long-distance travelling) as well as the saturation of the railways in the mid-1920s to the 1930s. The saturation of the railways (both in terms of infrastructure length and market importance) may have provided an "opportunity window" for the emergence of a competitor for long-distance travel in form of the car.

Roads

Extensive road networks existed long before the advent of the automobile. The French road network of the 1890s consisted of over 500,000 km of roads (Toutain, 1967), compared to 1985 with 800,000 km (Ann. Stat., 1987). The road infrastructure of the UK already extended over 177,000 miles in 1920 (Feinstein, 1988), compared to about 218,000 miles in 1986 (DOT, 1987). Even in the case of the USA the total road network did not really increase substantially. The earliest figures report some 3.16 million miles of roads in 1921, a figure which increased to 3.86 million miles in 1985 (US DOC, 1975 and US DOC, 1987). The road infrastructure was thus already developed to a considerable degree in the pre-automobile age. With an extensive road network already in existence, an obvious question which arises deals with the quality of the road network and of road transport in general, which, as discussed in Section 3.2 above, turned out to be inadequate to effectively compete with the railways.

With the possible exception of France, the quality of roads during the 18th century was extremely (even by contemporary standards) poor. However, road engineering improved substantially during the early 19th century, with important contributions by Thomas Telford (1757–1834) and John McAdam (1756–1836). The main engineering improvement was a new system of road foundations and water drainage, over which was laid compact, crushed stone which prevented water from reaching the road foundations. This meant that roads became resilient against surface water run-off, which had previously caused the rapid deterioration of road surfaces. As a result, road reconstruction, for instance in the form of *chaussées*, was carried out throughout Europe, resulting in a relatively high quality (in terms of road foundations, the surface of water-bound, *macadamized* roads, however, was not particularly smooth and certainly not dust free) infrastructure for horse carriages. Road construction was carried out by state authorities or by private turnpike trusts* (especially in England and the United States) and by 1870 the UK had about 160,000 miles of *macadamized* roads, France some 100,000 miles and Prussia some 56,000 miles (Forbes, 1975). Unfortunately, our quantitative information of the functional and quality characteristics of roads during the 19th century ends here. In fact the relatively few road statistics available all concentrate on an *institutional* (*administrative*) account of roads (i.e., disaggregation into national, provincial, or local roads, etc.); practically no information on the *functional* characteristics (width, type of surfacing, etc.,) is available**. We can only hope that in future, economic historians will shed some light on the question of 19th century road quality.

---

* A good account of turnpike trusts is given for the UK by Dyos and Aldcroft, 1969; Ginarlis and Pollard, 1988; and Voigt, 1965. For the USA see Taylor, 1962.

** Despite intensive literature research, no quantitative statistics for 19th century roads by surface type could be identified. The only noteworthy exception is the thorough study of Lepetit, 1983, for France; unfortunately only the period up to 1840 is covered. Lepetit reports that by 1836 some 3,800 km (11 percent) of the 33,400 km *routes royales* (i.e., what would be termed today "national roads") in France were paved.

Nevertheless, although we lack quantitative data on the quality (carrying capacity and transport speed allowed) of the road network during the 19th century, we can conclude that road infrastructure, at least in terms of the *size* of the road network, was already developed to a considerable degree, and was continually improved in terms of quality throughout the period of railways expansion. In the absence of a new technological (prime mover) base to replace the animal drawn carriages all efforts at improving road infrastructure proved unsuccessful in competing with railways. Voigt, 1965, estimates that the transport speed on the best road connections in France did not exceed 12 km/hour by the mid-19th century. As a result, railways transported practically all additional passengers and goods in all European countries throughout the 19th century and had an absolute market dominance in the transport sector by the end of the century. Road transport was finally confined to providing complimentary distributive services for the railroads, including in particular urban transport (horse tramways and omnibuses, *fiacres* and goods transport), with all the resulting problems of traffic jams, noise, and environmental pollution due to horse manure (Barker and Robbins, 1974).

The development of smooth, dust free road surfaces (an infrastructural precondition for the spread of the automobile) occurred considerably before the first automobiles appeared on the market. Paved roads (with stone or wood, as for instance in St. Petersburg) had existed in urban areas and in the vicinity thereof since the Middle Ages and had found widespread application in European cities in the 18th century. The first concrete roads (especially after the invention of Portland cement in 1824) were built in the USA as early as the 1850s. A second innovation was the asphalt road (the Place de la Concorde in Paris was paved with asphalt mastic as early as 1835, see Hamilton, 1975).

The existence of smooth road surfaces before the advent of the automobile can be best illustrated in the USA (US DOC, 1987). In 1904, 154,000 miles of roads were already surfaced for a meager car population of 55,000 (i.e., around 4,550 meters of road per car) compared to the current figures of 3.48 million miles of surfaced roads for a passenger car population of 135.7 million (i.e., 41.3 meters of road per car, US DOC, 1977). Thus, the available infrastructure in terms of surfaced roads for the first cars was over a factor of 100 larger than for the present car dominated transport system in the USA. This is an indication that road infrastructure development significantly *preceded* the diffusion of the automobile.

A quantitative analysis of the development of surfaced road infrastructure for the 20th century faces similar data availability problems as for the 19th century. Continuous time series data are available only* for the USA and the USSR and are analyzed in Figure 3.3.1. An interesting fact emerges. Despite a completely different situation with respect to private car ownership in the two countries, the diffusion of the surfaced road network proceeds at the same rate (Δt of 64 and 66 years), the USSR lagging about 30 years behind the USA. Again as with other transport infrastructures (canals, railroads) the ultimate saturation levels of diffusion are different in the two countries, reflecting different (geographical, economic, etc.) boundary conditions for the development of the surfaced road network. Ultimate saturation levels for the surfaced road network are estimated** to be 6.0 and 1.4 million km in the USA and the USSR respectively.

The estimated ultimate length of surfaced roads for the USA is 6 million km (3.75 million miles) compared with the 1985 figure of 5.6 million km. The length of the surfaced road network appears to have evolved along a logistic trajectory as shown in Figure 3.3.1. As the *total* road network stayed almost constant (it increased from 5 to 6.2 million km between 1920 and 1985) we can conclude that the diffusion of surfaced roads was essentially a substitution process, where surfaced road mileage replaced unsurfaced mileage; an observation made by Nakicenovic, 1986. The diffusion of the surfaced road network in the USA is currently almost complete (in 1985 surfaced mileage was 93 percent of the estimated K). On the other hand, as we will discuss below, the diffusion of the automobile is still far from being saturated. Car diffusion in any case proceeded both at a slower rate (longer Δt) and with a considerable time lag (lagged t_0) *after* the diffusion of the surfaced road network. This further corroborates our earlier conclusion that the car followed the development of (surfaced) road infrastructure and not vice versa. This does not contradict the fact that with increasing car densities significant *upgrading* (increasing throughput by multiple lanes or some network additions in the form of long-distance connections, i.e., interstate highways or *Autobahnen*) of the existing road network became necessary.

---

* Data on the percentage of different type of surface coverage on German federal roads (*Reichsstraßen*) for 1913, 1925, and 1939 are reported in *Der Elsner*, 1985. Some data on the length of surfaced roads in Germany in the 1930s are also contained in Voss, 1947 and Wienecke, 1956, however, these constitute too few observations to permit a quantitative analysis. These figures suggest that by the mid-1920s some 14 percent of *Reichsstraßen* were already surfaced. This figure increased to about 50 percent in the mid-1930s. By that date, around 25 percent of all major roads (*Reichs-* and *Landstraßen*) were surfaced. Estimates for the UK (Feinstein, 1988) indicate that by 1920 some 7 percent of all roads were surfaced, a percentage which increased to 24 percent by the mid-1930s, i.e., a figure similar to Germany. By 1920 the number of cars accounted for only between 0.2 (Germany) and 1 percent (UK) of present day figures.

** As the diffusion processes are not yet completed, statistical measures of uncertainty are provided. We estimate that there is a 90 percent probability that the ultimate length of surfaced roads will be between 5.6 to 6.3 million km in the USA and between 1.2 to 2.3 million km in the USSR (the higher uncertainty for the USSR is because the diffusion process has progressed to only slightly more than 50% of K). R^2 values of the estimates are 0.9923 and 0.9838 for the USA and USSR, respectively.

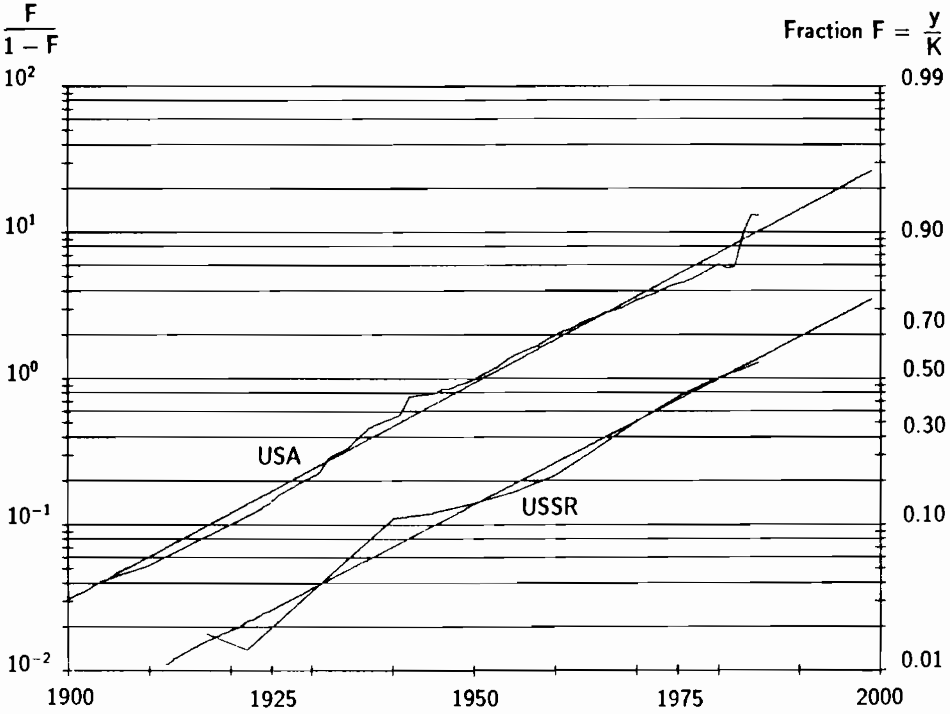


Figure 3.3.1. Length of surfaced roads in the USA [K 6.04 10^6 km; t_0 1951; Δt 64 years] and USSR [K 1.43 10^6 km; t_0 1980; Δt 66 years]. (Data source: US DOC, 1975, 1980, and 1987; Narod. Khoz., 1982, and Kommunist, 1987.)

The diffusion of surfaced roads in the USSR to the present level of 812,000 km (around 56 percent of our estimated K) is reasonably described by the logistic diffusion process in Figure 3.3.1. The process appears rather turbulent in the earlier history of the USSR, with the effects (decrease between 1917 and 1922) of the October Revolution and the ensuing civil war, as well as the very rapid increase up to 1940 (similar to the "overshooting" of railway construction in the same time period) being clearly visible. A regular diffusion pattern was reassumed only in the 1950s. The dynamics of this diffusion process are identical to that of the USA, which is not really surprising, considering the importance of road transport in the Soviet Union. Presently road transport (although not through individual cars but by public buses) is the most important means of long-distance passenger transport in the USSR, and in 1986 it accounted for 44 percent of intercity passenger-km (Narod. Khoz, 1986). This compares to around 80 percent in the USA (US DOC, 1987). This somewhat similar *functional structure* in the modal split between the two countries will be taken up again in Chapter 4.

Diffusion of the automobile at the global level

After having considered the diffusion of the appropriate infrastructure, let us now consider automobile diffusion in a number of countries. First, we will define what we consider an automobile (car) to be. The first cars were very different from the predominant technological design of present day cars. Early cars may instead be characterized as being "horseless carriages", often with open (and thus rather uncomfortable) bodies; construction material (wood as the structural material), axels, interior design, such as seats, steering design, etc., were all taken directly from horse carriage designs or were derivates. This provides a clear picture of the market niche in which early automobiles were competing: The automobile emerged and grew by replacing horse drawn vehicles.

Even the technological design of the major difference from horse carriages, i.e., the new power source of automobile traction energy (fossil energy), was far from being standardized. In fact, four technological routes competed as the prime mover of the automobile: steam, diesel, and gasoline engines and electric motors. It is one of the characteristics of the early phase in a technology's life cycle, that a large number of diverse technological designs* coexist and compete.

The final outcome of this competition is highly uncertain and essentially unpredictable. It depends on a large number of small, random like events, until the system finally "locks in" and a predominant technological design** emerges (as argued by Arthur, 1983, and 1988). Standardized technological design enabled mass production and significant cost reductions. The most prominent example is of course Ford's Model T, which reduced its selling price from US $850 in 1908 to as little as US $290 in 1926 (see in particular Abernathy, 1978, on technological change in the automotive industry). The technological and productivity (cost and performance) stage for the diffusion of the automobile and the replacement of the horse carriage was thus set.

Finally, one has to mention another precondition for the diffusion of the automobile: the overcoming of institutional and societal barriers. As an example we mention the abandonment of the 1836 "Red Flag Act" in the UK. This Act (see Voigt, 1965) required that, for all

---

* Here one should also mention the invention of the bicycle in road transport technology (1842). Although the bicycle did not significantly change the structure of road transport, it provided an important predecessor technology for the motorcycle, which was the most important (in terms of their number) motorized road vehicle in most European countries well into the 1950s.

** The automobile with the gasoline internal combustion engine. First examples include the motor vehicle with a gasoline engine constructed by Siegfried Marcus in Vienna 1875 (patented 1882) or the 1.5 horsepower (hp) automobile constructed by Gottlieb Daimler in 1886.

steam driven road vehicles, a person with a red flag and a bell had to walk 20 yards before the vehicle to warn others. Steam vehicles were also not allowed speeds exceeding 4 km/h. Yet another factor preparing the (social) grounds for the diffusion of the automobile was the increasing realization of the negative aspects resulting from the dense horse carriage traffic in urban areas with all the associated problems of congestion, energy (feed) supply, "garage" problems, high accident rates, noise, and environmental pollution stemming from horse manure (see Barker and Robbins, 1975, for an account of the road traffic situation of 19th century London).

Figures 3.3.2 and 3.3.3 give an account of the expansion of the automobile in terms of passenger cars registered worldwide (currently the total is close to 400 million)* and by broad geographical area. Figure 3.3.2 illustrates the expansion of the automobile in terms of passenger cars registered at the world level, as well as for the USA and the Organization for Economic Cooperation and Development (OECD) countries outside the USA.

Three important features are noticeable in Figure 3.3.2:

(1) Two growth phases are clearly revealed by the different slopes on the logarithmic scale. The first very rapid expansion phase lasted until the 1930s with an average exponential growth rate of around 30 percent per year; the second phase had a significantly slower growth rate of 5 percent per year on average from 1930 to 1985.

(2) Prior to World War II, automobile growth was greatest in the USA. The shaded areas of Figure 3.3.2 indicate car registrations in OECD countries outside the USA and outside OECD (i.e., the difference between OECD and the world total) and clearly reveal the absolute dominance of the USA in the first phase of the diffusion of the automobile. Up to 1930, the USA alone accounted for as much as 80 to 90 percent of all passenger cars registered worldwide. By 1950, the share of the USA had fallen to around 76 percent, the remainder of the OECD countries accounted for 16 percent and the rest of the world for around 8 percent of passenger

---

* Data from 1930 onwards are from MVMA, 1983, and 1985. In the absence of available statistics on total car registrations in OECD countries, we use a proxy sample of 12 countries: Australia, Austria, Canada, France, FRG, Italy, Japan, Spain, Sweden, New Zealand, UK, and USA. Between 1930 and 1950 our sample (constantly) accounts for over 90 percent of the world total. Consequently we have applied the same percentage to the period prior to 1930 to make an estimate of world passenger car registrations. Data sources are: Mitchell, 1980, and 1982; IRF, 1980, and 1986; as well as updates and consistency checks by national statistics. It should be noted, that official registration figures even in industrialized countries are affected by measurement errors. Alternative estimates (Polk, 1987) for instance in the case of the USA suggest that actual passenger cars in use may differ by as much as 10 percent from the officially recorded registration figures.

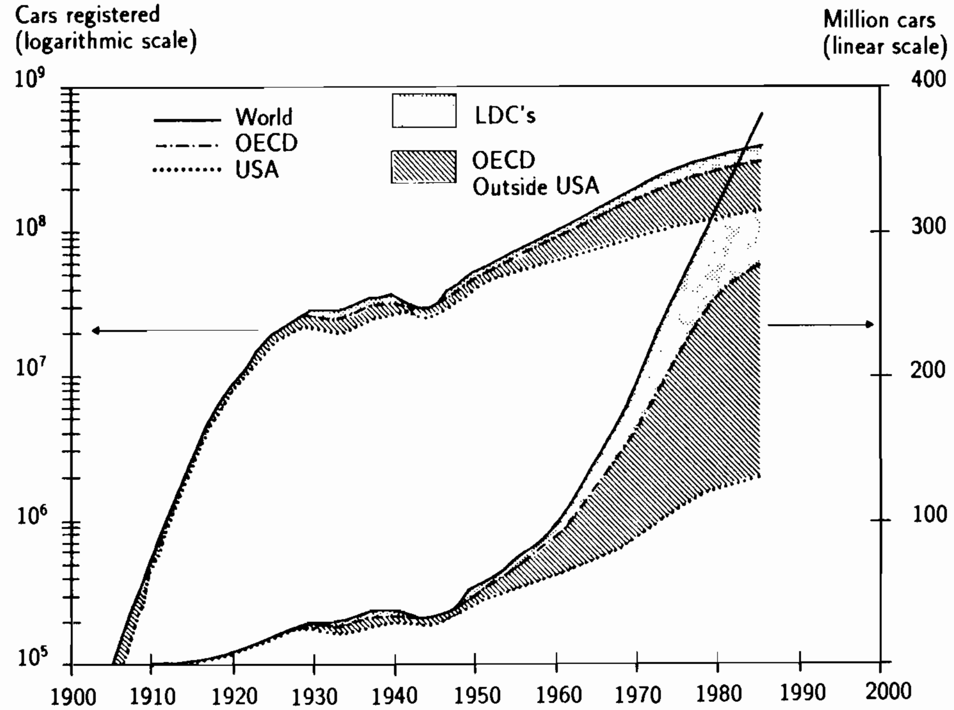


Figure 3.3.2. Passenger cars registered worldwide, linear and semi-logarithmic scale.

cars registered. Since 1950, the share of the USA has fallen to around 35 percent, while the remainder of OECD countries has risen to about 40 percent of the world total.

(3) The share of countries outside the OECD remained basically very small, below 10 percent of the world total up to 1960. Since 1960 this share has risen to about 25 percent.

In the first phase of diffusion between 1900–1930 the number of passenger cars registered worldwide increased by more than four orders of magnitude, to some 30 million passenger cars, of these over 23 million were registered in the USA. In the second period, passenger car registrations increased from around 30 million cars in 1930 to close to 400 million by 1985, with significant growth also occurring outside the USA. Thus, we conjecture, that the diffusion of the automobile is characterized by two periods. The spectacular growth of the automobile in the first period up to 1930 is the result of the diffusion of the car into a specific market niche (road transport), which was previously held by horse driven carriages and where its comparative advantage was particularly decisive. Consequently, the substitution of horses in the transport sector proceeded fast and became almost complete in all

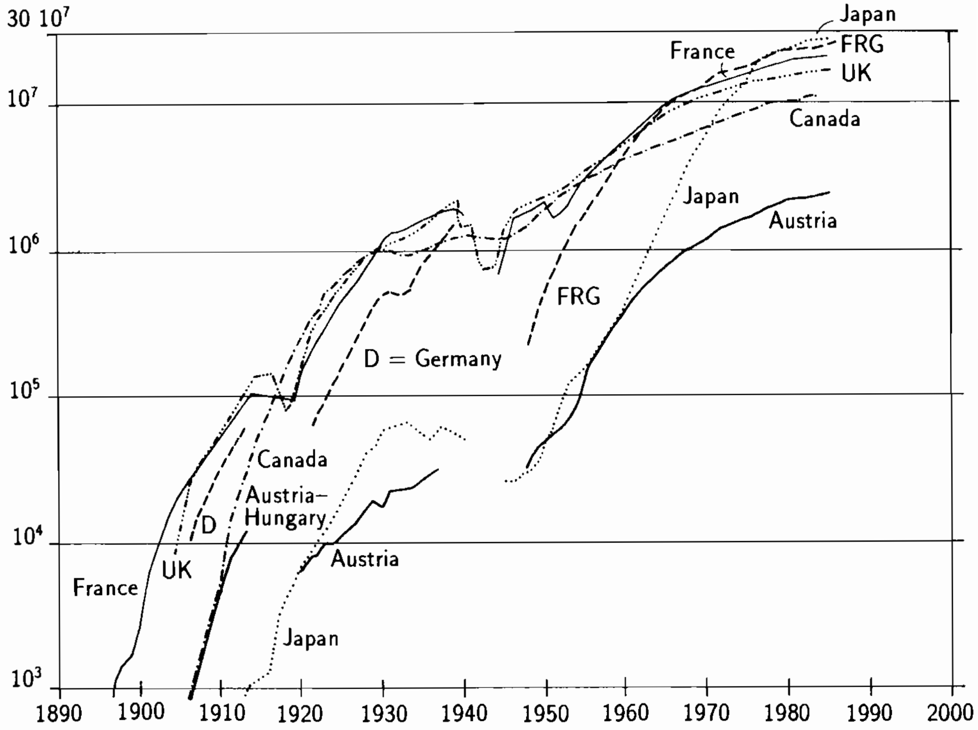


Figure 3.3.3. Growth of passenger cars registered for selected industrialized countries.

countries by the 1930s, as illustrated below. The second growth period is characterized by diffusion into new market niches, either previously held by railways, or in newly emerging applications such as leisure and weekend travel, commuting, etc.

The two distinct phases in the growth of the automobile discussed above, is further confirmed if we look at the growth of car registrations in a number of OECD countries outside the USA (see Figure 3.3.3). Although the picture is more turbulent due to the effects of the two World Wars and the consequent rapid catch up of a number of countries (Austria, FRG, and Japan) one can see that the 1930s constitute an important "watershed" period in the diffusion of the automobile in these countries too. Up to 1930 the number of cars registered increased by more than three orders of magnitude, compared to approximately one order of magnitude in the 50 years that follow. Again we conjecture that this is the result of the diffusion into different market niches in the two periods (road transport, held by horse carriages in the first, and the transport sector in general in the second). Figure 3.3.4 analyzes the diffusion of the total number of passenger cars registered worldwide consistent with our hypothesis of two diffusion periods.

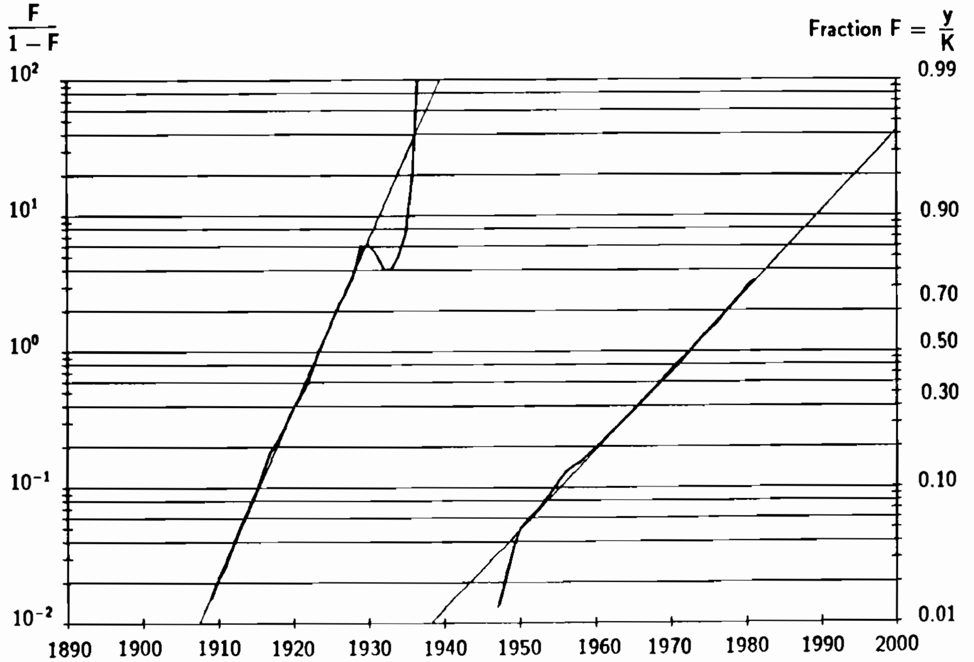


Figure 3.3.4. Two diffusion pulses in the expansion of the number of passenger cars registered worldwide [First pulse: K 34.6 million cars; t_0 1923; Δt 15.4 years. Second pulse K 463.9 million cars (including intercept); t_0 1974; Δt 36.2 years]. (Data source: WVMA, 1983, and 1987, for period prior to 1930 own estimates).

The data appear to support our working hypothesis, that the expansion of the world automobile fleet indeed proceeded along the two logistic diffusion pulses depicted in Figure 3.3.4. The first phase in the diffusion of the automobile is characterized by a regular logistic diffusion process, which proceeded rather fast (Δt of 15.4 years) in order to reach a saturation level of 34.6 million cars by 1937. The downward deviation occurring after 1930 is the effect of the Great Depression, with a decrease in the number of cars registered in the USA and Canada. This deviation was, however, absorbed again by 1936.

The second phase in the diffusion of the automobile worldwide occurs after 1950 (by which date the effects of World War II had been overcome). The expansion of the world automobile fleet appears to proceed smoothly along the logistic diffusion pattern shown in Figure 3.3.4. An interesting observation about Figure 3.3.4 is that the post-1973 developments in the oil market did not appear to have had a noticeable effect on the diffusion process. This second expansion phase of the automobile is characterized by a time constant twice as large as the first diffusion pulse (Δt of 36.2 years compared to 15.4 years) and saturation, if the diffusion pattern continues to unfold as in the past, would occur around the year 2010 at a level of around 464 million cars.*

The growth of worldwide car registrations appears thus to be approaching saturation, with between 27 to 44 percent growth potential remaining, taking best fit and high uncertainty band estimates respectively. This would point at the progressive exhaustion of the economic growth stimulating effects of the expanding automobile industry, characteristic for the "oil and car" age. Before we look at the diffusion of the automobile in a number of countries we corroborate our hypothesis that there are two phases in the diffusion of the automobile.

The replacement of horses

The impact of the diffusion of the automobile on road transport at the beginning of this century will be analyzed using three different measures. We will analyze first the growth in the *number* of road vehicles (horse carriages, cars and motorcycles) in the UK, then study the evolution of the *performance* of road transport by horses and the automobile (in terms of share in the passenger-km and ton-km) in France, and conclude with an analysis of the growth in the *number of prime movers* in road transport (number of horses versus number of cars) for the USA. These three different measures are taken to support our hypothesis that the first phase of the diffusion of the automobile consisted indeed of the replacement of horses by cars.

Figure 3.3.5A reports the share of motor driven road vehicles in the total private passenger vehicle fleet in the UK. Figure 3.3.5B extends this analysis to a more systematic account of the diffusion of motor vehicles in different market segments (the symmetrical displacement data and substitution curves of horse drawn road vehicles are omitted in Figure 3.3.5B). The analysis of the substitution processes is based on data recently estimated by Feinstein, 1988 (referring to Thompson, 1976), on the number of horse drawn road vehicles (non-farm carts, vans and wagons, horse omnibuses and stage coaches, hackneys, and 2- and 4-wheel private carriages) compared to motor road vehicles (private cars and taxis, buses, motorcycles and trucks)

---

* We estimate that there is a 90 percent probability that the saturation level of the diffusion process will be between 404 to 524 million cars (1984: 364.8 million). This estimate includes the intercept (saturation level of previous growth pulse of 34.6 million cars). R^2 of the estimate is 0.9998.

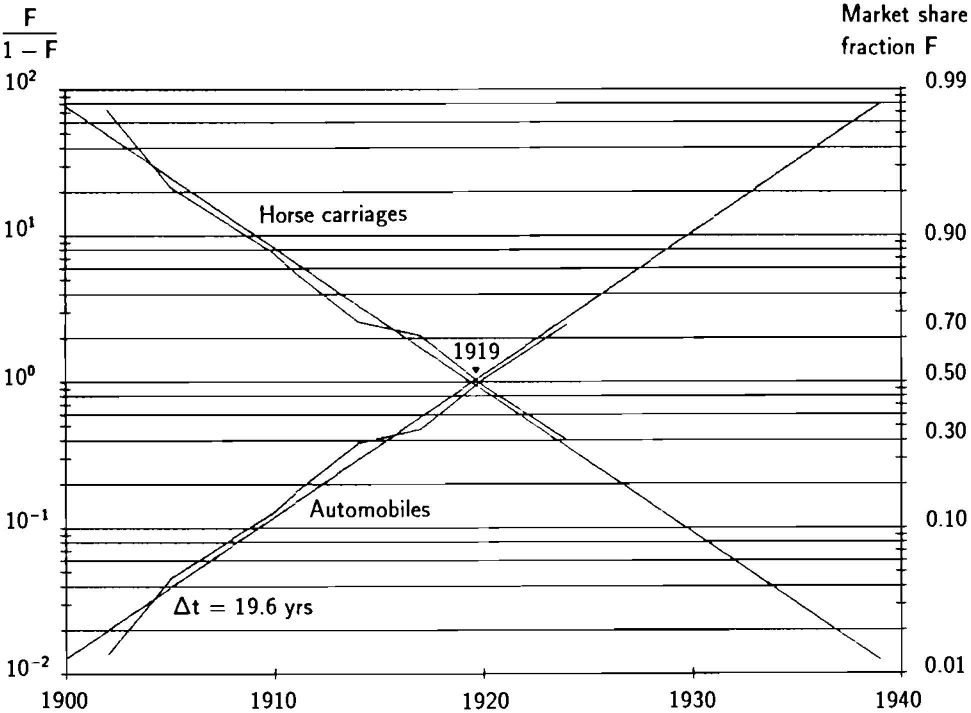


Figure 3.3.5A. Substitution of hackneys and carriages by motor vehicles (cars, taxis, and motorcycles) in private passenger vehicle fleet in the UK [t_0 1912; Δt 16 years]. (Data source: Feinstein, 1988, Mitchell and Deane, 1971, and Thompson, 1976.)

as reported in Mitchell and Deane, 1971. An interesting insight of this analysis of a historical substitution pattern is the fact that the automobile did definitively not diffuse into a vacuum. By 1910 over 1.3 million horse drawn vehicles existed in the UK (compared to some 170,000 motor vehicles), i.e., about one horse carriage per 30 inhabitants. The number of horse drawn road vehicles decreased to 530,000 (including 330,000 non-farm carts and wagons) by 1924. Thus, we can conclude that the first phase of automobile diffusion indeed was a substitution process of horse drawn vehicles in road transport.

This process was rather fast (Δt of 19.6 years; t_0 1919), but proceeded at a different pace for various market segments. The fastest replacement occurred in the area of mass public transport, where horse omnibuses and stagecoaches were replaced with a Δt of only 8.1 years (t_0 1912), followed by private road vehicles (hackneys and 4- and 2-wheel private carriages were replaced by private cars, taxis, and motorcycles) with a Δt of 16.1 years (t_0 1912). Replacement in the area of goods transport (the lowest value market niche for the first automobiles) proceeded with a Δt of 22.5 years slower and occurred

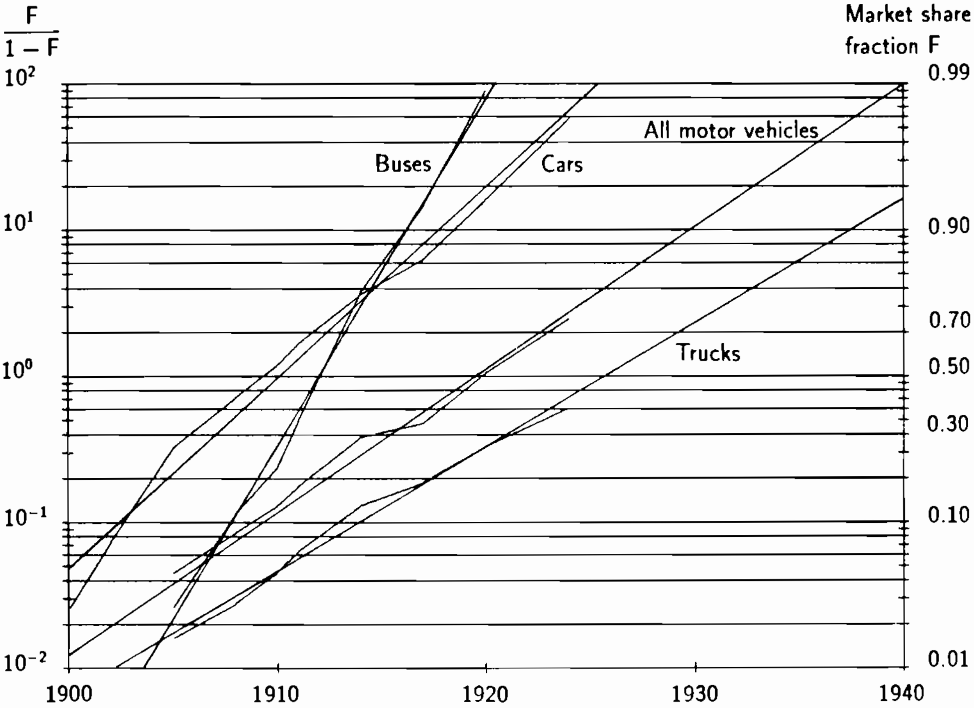


Figure 3.3.5B. Share of motor in road vehicles [Mass public transport (buses): t_0 1912; Δt 8 years. Private passenger vehicles (cars): t_0 1912; Δt 16 years. Goods transport vehicles (trucks): t_0 1926; Δt 22.5 years. Total road vehicles (all motor vehicles): t_0 1919; Δt 19.6 years]. (Data source: see Figure 3.3.5A.)

also at a later date (t_0 1926). Still, we can conclude that by the end of the 1930s the horse had virtually disappeared as a means of providing road transport services in the UK. Within a relatively short time period, between 1900 and 1930, we observe a complete market takeover of the automobile in road transport.

Our analysis for the UK is further confirmed when we consider the replacement of the horse in road transport in France, as done in Figure 3.3.6. Again, the symmetrical displacement data and negative diffusion curves of horse vehicles are not plotted in the figure for reasons of clarity. Thanks to the research of Toutain, 1967, we can assess the market impact of the automobile in measuring the relative market share of horse and motor vehicles in road transport output (passenger- and ton-km). The automobile first penetrated into road transport by replacing horse drawn vehicles. The dynamics of this process in France are quite similar to the UK. The displacement of horse carriages for passenger transport proceeded with a Δt of 15.3 years (t_0 1909) somewhat faster than in the UK and occurring also at a slightly earlier date. As in the UK, the diffusion of the automobile into lower

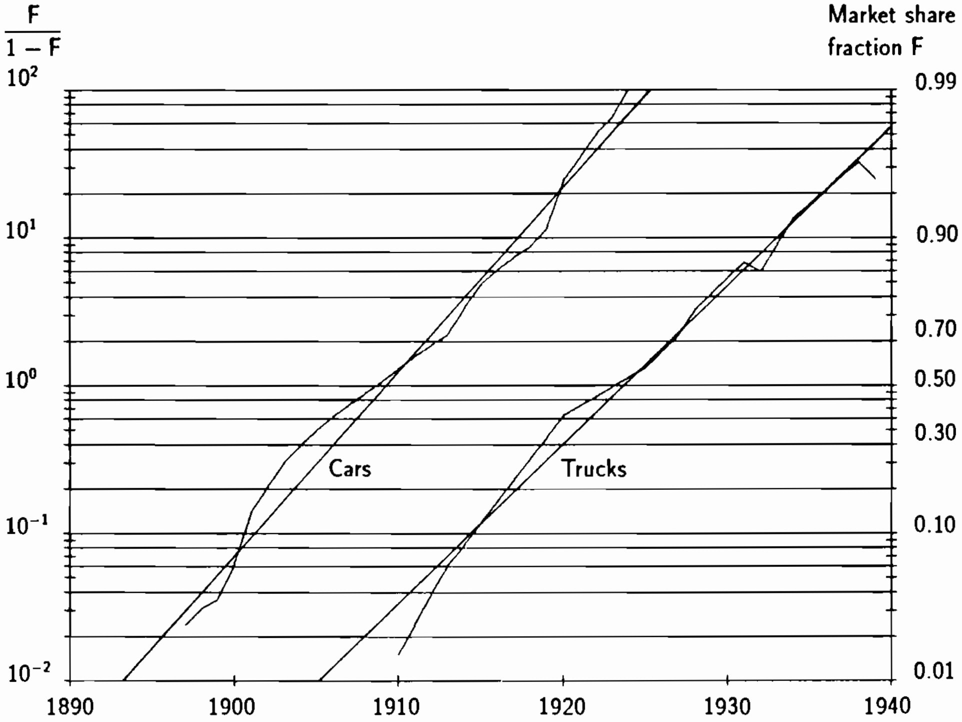


Figure 3.3.6. Share of motor vehicles in output of road transport in France, measured by passenger- and ton-km performed [Passenger transport substitution: t_0 1909; Δt 15.3 years. Goods transport substitution: t_0 1924; Δt 17.7 years]. (Data source: Toutain, 1967.)

value market niches (goods transport) was slower and occurred at a later date (Δt of 17.7 years and t_0 1924). Similarly, the automobile had effectively displaced horses from the road transport market by the end of the 1930s. Our estimates for France allow us also to assess the impact on market volume expansion, due to the diffusion of the automobile.* Road passenger transport, which was stagnant at a level of around one billion (10^9) passenger-km before the appearance of cars and motorcycles, increased by a factor of 40 until the end of the 1930s. The output level of goods transported by horses, which had remained at a level of 3.3 billion ton-km before the advent of the automobile, increased to over 12 billion ton-km (i.e., by a factor of nearly four) up to the end of the 1930s. Consistent with our earlier arguments the pace of diffusion as well as the impact on market volume of the

---

* Recall here the discussion of the technological diffusion model of Metcalfe, 1983, in Chapter 2 above. This particular model argues that the diffusion of a new technique (replacing existing ones) results in a market expansion "impulse" to a new (equilibrium) level. For France the market expansion impact of the diffusion of the automobile in replacing horses is substantial.

diffusion of a new technology appears significantly higher in high-value (price insensitive) market niches such as passenger transport.

We conjecture that the situation in other countries closely follows the pattern as outlined above for the UK and France. Although we do not have comparable data for Germany, we can nevertheless provide some indication on the *employment effect* of the diffusion of the automobile. Voigt, 1965, presents contemporary statistics of the German Reich on people employed in road transport. In 1908 about 195,000 coachmen compared to about 5,000 car and truck drivers were employed in Germany. This ratio changed drastically to around 110,000 coachmen and over 200,000 car and truck drivers by 1933. Thus, whilst nearly 100,000 coachmen lost their job or had to be retrained as car drivers, this loss was more than compensated for by new jobs created in automobile transportation. The increase in market volume brought about by the diffusion of automobile technology, resulted thus in a net increase of over 50 percent in the number of driving personnel in the period from 1908 to 1933.

Finally, let us analyze the first phase of the diffusion of the automobile in the USA, in the country with the highest car density in the world. Credit has to be given to the first analysis of the replacement of horses as a source of transport motive power, carried out by Nakicenovic, 1986. In analyzing the number of horses and mules used for transport purposes (i.e., excluding farm animals), the picture that emerges is similar to that of the UK and France. Despite the relatively crude measure for the substitution process allowed by the available historical statistics (one car could in fact have replaced more than one horse, since carriages were often pulled by more than one horse), a regular substitution process is shown in Figures 3.3.5 and 3.3.6. With a Δt of 12.3 years and t_0 in 1916, the substitution process was somewhat faster than in Europe. The results for the USA confirm our earlier conclusion that by 1930 the horse had practically disappeared as a means of road transport in all industrialized countries.

In Figure 3.3.7 the substitution process of horses by automobiles is illustrated by plotting their respective absolute numbers and the estimated numbers based on a logistic substitution model. This substitution resulted in a decrease in the number of draft animals used for transport, i.e., from near to 4 million in 1910 to less than 400,000 by the end of the 1930s with a corresponding increase in the number of cars from 400,000 to over 20 million. By the time this substitution process was completed at the beginning of the 1930s, the US automobile density was close to 200 passenger cars per 1,000 inhabitants. This is a value comparable to the car density of present day Japan (240 cars per 1,000 inhabitants in 1987). This is another illustration of the specific circumstances explaining the present high automobile

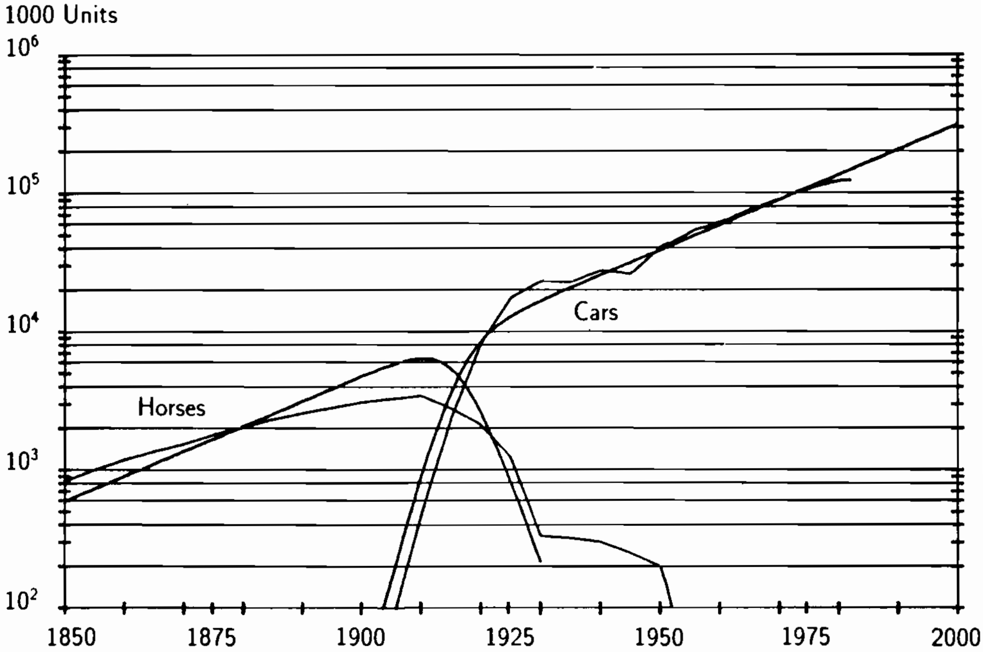


Figure 3.3.7. Number of draft animals (horses and mules) used for transport and number of cars in the USA (empirical data and model estimates from a logistic substitution model). (Source: Nakicenovic, 1986.)

density in the USA, which makes it all the more unrealistic to use the US situation as a guide for developments in other countries. In addition the unique initial starting conditions for the automobile age in the USA have to be considered. For instance, before the appearance of the automobile, the number of riding horses and mules in the USA in 1900 exceeded 40 per 1,000 inhabitants, which exceeds today's car ownership rate in Hong Kong, and is comparable to the current average car density worldwide (outside the USA) of around 50 passenger cars per 1,000 inhabitants.

Another observation, which can be made on the basis of Figure 3.3.7 is that the *total number of road vehicles* (i.e., horses plus cars) appears to have evolved along an exponential secular trend with an average growth rate of above 4 percent per year ever since 1850. This suggests an analysis of the diffusion of all road vehicles, irrespective of the replacement process between horses and cars occurring inside the total road vehicle population. This analysis is reported in Figure 3.3.8. If the growth of the total number of road vehicles in the USA is analyzed under the hypothesis of a single diffusion process, it appears that the growth of the number of road vehicles proceeds after initial

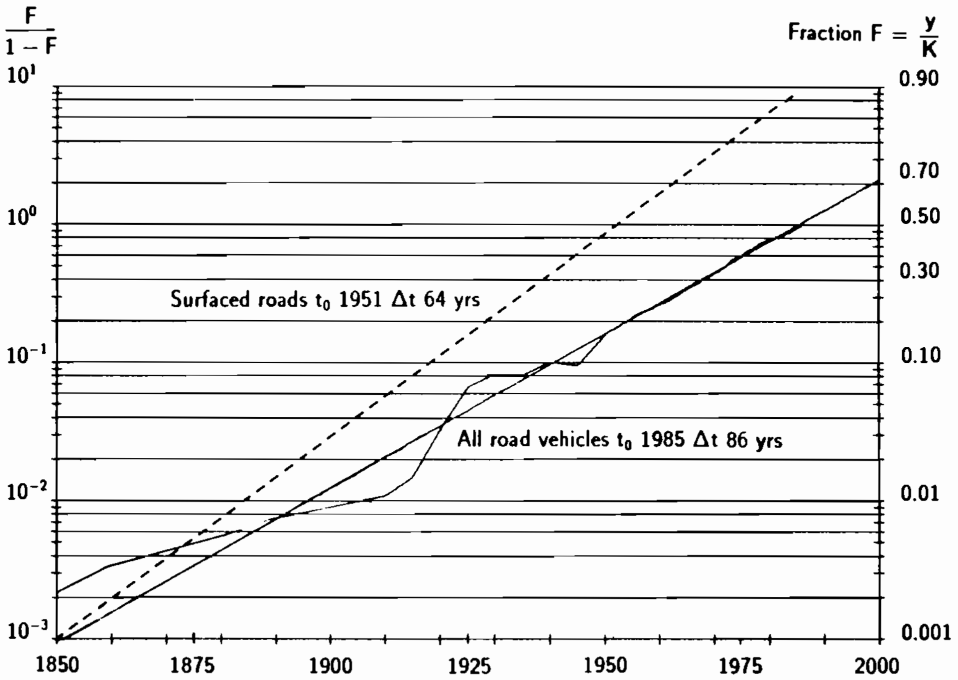


Figure 3.3.8. Growth in total number of road vehicles in the USA [K 352.6 million vehicles; t_0 1985; Δt 86 years] and growth of surfaced roads (from Figure 3.1.1). [Source: Updated (US DOC, 1987) from Nakicenovic, 1986.]

significant turbulences* (but still along the long-term diffusion pattern) finally "locking in" in a regular growth path along the logistic diffusion trajectory. The implications of this process, should it continue its historical trend, are particularly noteworthy with respect to the estimated saturation level of around 350 million vehicles, which would correspond to a doubling of present day figures and a resulting road vehicle density of over one vehicle per inhabitant. Although such a conjecture appears unrealistic, based on today's organizational patterns of car ownership and usage, there are some tourist centers where car densities exceeding one car per inhabitant can already be observed today. Could it be that in the future such large numbers of additional road vehicles will be used in the form of rented cars or similar

---

* In order to report the whole period in history for which data were available, we also plot in Figure 3.3.8 the phase below 1 percent of the ultimate saturation level, which is usually omitted in the parameter estimation and graphic representation of results. Note also that the ordinate in Figure 3.3.8 stops at 90 percent of K. As the diffusion process has presently progressed to only 50 percent of the estimated final saturation level, the statistical uncertainty involved in parameter estimation is substantial. Assuming a 10 percent data error (to cover the different numerical convergence of various estimation algorithms), we estimate a 90 percent probability that K (352.6 million vehicles) will be between 207 to 516 million vehicles (1986: 176.5 million). R^2 of estimate is 0.9966.

organizational schemes to provide local and regional transport, with long-distance transport being provided for by other modes of transportation (e.g., aircraft or Maglevs)?

Only the future will tell, whether such a scenario may be realistic. Let us return instead to our previous statement, that the development of road infrastructure (surfaced roads) preceded the diffusion of road vehicles. In Figure 3.3.8 we have also plotted the diffusion of surfaced roads in the USA (from Figure 3.1.1) to provide final empirical evidence supporting our conclusion that the diffusion of the automobile *followed* (and not enforced) the diffusion of road infrastructure. The diffusion of surfaced roads in the USA (t_0 1951) precedes the diffusion of road vehicles (t_0 1985) by over 30 years and progressed with a Δt of 64 years faster than the expansion of road vehicles (Δt 86 years). Presently the expansion of surfaced roads appears to be practically complete, whereas possibly only half of the ultimate number of road vehicles has been reached. It is perhaps not incidental that the origin of the two diffusion pulses coincides in the 1850s, emphasizing the close relationship between infrastructure and the transport technology using it. Roads were first surfaced to serve horse driven vehicles and the same infrastructure could subsequently be used by the automobile, when in turn it started to replace the horse on the street.

We would like now to return to the beginning of the automobile age, and put forward some tentative propositions as regards the driving force behind the substitution process of horses by cars. We have concluded that the replacement of horses was practically complete in all industrialized countries by the 1930s, independent of which specific indicator was used to describe this substitution process. The speed of displacement of horses in road transport was rather fast, with Δts ranging from 12 (USA) to 20 (UK) years, meaning that within a period of one to two decades 80 percent of the stock of horses (and carriages) in the road transport sector were replaced.

What are the driving forces behind such a rapid process of technological change? Clearly, a complex vector of comparative advantages of the automobile over horses. Among these comparative advantages we rank in first place *improved performance* in terms of speed, range, and ease of use. Average transport speeds of automobiles increased quickly from around 30 to some 50 km/h [average transport speed on highways was higher; for 1940 an average speed of 40 miles (i.e., 60 km/h) is reported, (US DOC, 1975)], compared with the typical 10 km/h for average horse carriages. Higher speed and ease of use (no changing of horses over long-distance journeys) meant a considerable extension in the travel range which could be covered. Another aspect of ease of use of the automobile was a significant reduction in inventories [stables, storage of food – a horse consumes over 25 kg of food per day in the form of bulky hay, oats, etc., (Montroll and Badger, 1974)].

Finally, compared to the horse, the automobile can only be described as environmentally benign, at least in urban areas of dense traffic (Ausubel, 1989). Cars made less noise and also less emissions than the previous dense horse population, as long as the numbers of cars did not significantly exceed the previous horse population in cities. Although we now know that the quality (toxicity) of emissions of automobiles and horses is entirely different, the automobile nevertheless offered relief from one of the most urgent environmental pollution problems of urban areas around the turn of the century: horse manure. Montroll and Badger, 1974, estimate that a horse produces about 1,000 grams of excrement per mile (g/mile) travelled (635 g/mile solid and 300 g/mile liquid). This compares to 1980 piston engine standards in the USA of around 5 g/mile (CO, NO_x and hydrocarbons). Thousands of so-called "crossing sweepers" were employed in London who would clean a path before anyone crossed the street (at a cost of one halfpenny). Today this may sound anecdotal (provided by Montroll and Badger, 1974) but horse traffic did previously constitute a serious urban environmental problem.

We conclude our discussion on the first phase of the diffusion of the automobile by saying that it was (a) characterized by a process of substituting horse drawn vehicles, and (b) that this first phase of the diffusion process was essentially completed by the 1930s. Technological improvements in automobile design and further improved performance and comfort of automobiles has prepared the ground for the second phase of the diffusion of the automobile, i.e., its emergence as a mass transport mode and the beginning of competition with traditional long-distance carriers, in particular the railroads.

The second phase of automobile diffusion

The diffusion of private car ownership will be discussed below under two main assumptions. First, empirical evidence suggests that the diffusion pattern closely follows a logistic trajectory.* Our second assumption is based on the conclusions of our analysis of the first phase of the diffusion of the car. Since it was not until the end of the 1930s that horses disappeared as a means of road transport, we can only consider the number of cars as an appropriate indicator for passenger road vehicles after that date. Consequently (and by further

---

* Sensitivity analysis with alternative diffusion models yielded unsatisfactory results: bad statistical fit and also unrealistic (and between different models, widely diverging) saturation levels.

ignoring the disturbances resulting from World War II) we will analyze the second phase of the diffusion of the automobile by using data from the period 1950 to 1986 only. The analysis will use both absolute and relative measures, one being the number of cars registered, the other one being the number of cars registered per 1,000 inhabitants.

We have shown in the introduction to this Section that private car motorization has been primarily a phenomenon of OECD countries, which presently account for about three-quarters of all passenger cars registered worldwide. Consequently we will begin our analysis of post World War II automobile diffusion by examining a sample of 12 OECD countries, representative of the automobile "diffusion bandwagon". Subsequently, we extend the analysis to centrally planned economies and developing countries primarily for two reasons:

- to examine the likelihood of a similar scale of motorization in those countries as observed for OECD countries (concluding that similar automobile density levels appear unlikely to emerge in the future).
- we perform a "bottom-up" consistency check to examine if the projected world total is in agreement with the aggregated individual diffusion forecasts of our sample of 20 countries, which account for over 90 percent of world passenger car registrations (concluding that top-down and bottom-up diffusion approaches yield consistent results).

OECD countries

Table 3.3.1 presents the estimates of the diffusion of passenger cars in a number of OECD countries, as well as the total for the world (including centrally planned and developing countries). We show the timing (inflection point t_0, i.e., at $K/2$) and the duration of the diffusion process (Δt), the estimated ultimate level of saturation K, together with the corresponding uncertainty bands for the estimates (at 90 percent probability level). The actual car density at present (registered passenger cars) is also given. As can be seen from these figures, automobile diffusion appears close to saturation in most of the OECD countries and at the global level. In fact, in a few cases the lower uncertainty band has already been surpassed by actual registration figures. This is the reason why only best fit and high uncertainty band estimates are considered as ranges for the estimated saturation level K in the subsequent discussion.

Table 3.3.1. Passenger car diffusion,1) from 1950 to 1987 (in million passenger cars registered).

Country	t_0	Δt	Passenger Cars (million)			Present3) in % of K	R^2 of estimate
			K	K-range2)	Present (1986–87)		
Australia	1972	43.6	8.6	7.4–9.7	6.8	70–80	0.9996
Austria	1972	27.6	2.9	2.6–3.1	2.7	85–94	0.9978
Canada	1972	51.8	15.0	12.9–17.1	11.5	67–77	0.9969
France	1969	33.8	23.8	21.4–26.1	22.0	84–92	0.9958
FRG	1971	28.9	29.7	26.7–32.7	28.3	87–95	0.9949
Italy	1973	25.7	24.6	22.1–27.0	23.3	87–95	0.9924
Japan	1975	16.6	30.1	27.1–33.1	29.5	89–98	0.9993
Spain	1976	21.2	11.1	10.0–12.2	10.3	85–93	0.9989
Sweden	1965	32.4	3.4	3.1–3.7	3.4	92–100	0.9925
New Zealand	1970	47.7	1.9	1.7–2.1	1.6	76–86	0.9965
UK	1966	33.3	18.1	16.4–19.9	17.0	85–94	0.9977
USA4)	1970	59.74)	157.7	141.4–174.0	135.7	78–86	0.9985
World5)	1974	51.65)	463.9	403.6–524.3	364.8	70–79	0.9998

1) Data period used for parameter estimation: 1950–1986/1987.
2) With 90 percent probability.
3) Range corresponds to high estimate and best fit case respectively.
4) Two growth pulses ($\Delta t = 14.4 + 45.3$ years).
5) Two growth pulses ($\Delta t = 15.4 + 36.2$ years, see Figure 3.3.4). Latest data available: 1984. World total includes centrally planned economies.

(Data sources: Mitchell, 1980, and 1982; IRF, 1980, and 1986; and national statistics.)

A number of observations can be made from Table 3.3.1. First, the growth of passenger cars registered appears to be close to saturation in all the countries of our data sample. The remaining growth potential is rather small for all OECD countries analyzed, whereas for the world a figure of up to 30 percent higher than at present could, based on our model, be expected. A second observation deals with the very large heterogeneity in both the diffusion speed (Δts ranging from 17 years in Japan to over 50 years for Canada and the USA) and in the remaining growth potential. Third, the diffusion rates (Δts) seem to accelerate the later the diffusion process began.

Both the diffusion rate and the level of saturation thus appear to be influenced by the time the diffusion process was initiated (measured by the year when one percent of the saturation level was reached, i.e., t_0 minus Δt). Early starters grow slower and achieve higher ultimate density levels. But late starters have higher diffusion rates and therefore tend to catch up with the leaders, albeit reaching lower saturation levels. This is exactly the same phenomena as seen for the spread of railways. For instance, the automobile diffusion process was initiated in Canada and the USA around the turn of the century, grew slowly (Δt around 50 years) and results in very high car density levels

Table 3.3.2. Passenger car diffusion1) from 1950 to 1987 (in passenger cars registered per 1,000 inhabitants).

Country	t_0	Δt	Cars per 1,000 inhabitants				
			K	K-range2)	Present (1986–87)	Present3) in % of K	R^2 of estimate
Australia	1965	46.7	495.6	441.6–549.5	431.6	79–87	0.9990
Austria	1971	27.2	366.4	330.0–402.8	336	83–92	0.9975
Canada	1968	69.6	615.4	526.4–704.4	454	65–74	0.9924
France	1967	34.5	422.5	380.6–464.4	394	85–93	0.9953
FRG	1971	30.7	493.1	443.8–542.4	463	85–94	0.9941
Italy	1972	26.1	426.5	384.0–469.0	408	87–96	0.9920
Japan	1974	16.6	245.2	221.2–269.1	241	90–98	0.9991
Spain	1976	21.9	283.9	255.5–312.2	265.7	85–94	0.9988
Sweden	1964	32.8	401.8	362.0–440.8	401.0	91–100	0.9950
New Zealand	1967	60.7	581.0	508.1–654.0	487.5	75–84	0.9946
UK	1966	36.5	333.3	300.2–366.3	318.0	87–95	0.9895
USA4)	1970	57.84)	611.0	568.0–654.0	562.0	86–92	0.9950
World5)	1970	34.15)	62.7	58.0–67.3	61.0	91–97	0.9950

1) Data period used for parameter estimation: 1950–1986/1987.

2) With 90 percent probability.

3) Range corresponds to high estimate and best fit case respectively.

4) Two growth pulses (Δt = 14.5 + 43.3 years).

5) Two growth pulses (Δt = 14.2 + 19.9 years). Latest data available: 1984. World total includes centrally planned economies.

(Data sources: see Table 3.3.1.)

as shown in Table 3.3.2. In Japan the diffusion process started some 60 years later, but due to higher diffusion rates grew faster (caught up with a Δt of some 17 years), and appears to be saturating at the same time as the USA and Canada but at a lower level. Table 3.3.2 complements the analysis reported in Table 3.3.1 by using relative car density (passenger cars registered per 1,000 inhabitants) as an additional indicator.

In comparing Tables 3.3.1 and 3.3.2, one can conclude, that the results of diffusion analysis using both absolute and relative measures are consistent. Differences in the estimated Δts for Canada, New Zealand, and the world total are the result of strong divergences between population growth and passenger car registration growth rates. However, the analysis of both indicators agrees with the conclusion about the limited growth potential remaining for further automobile diffusion. We again point at the strong heterogeneity in the ultimate saturation density levels that emerges from our analysis. Early starters such as the USA and Canada may reach saturation density levels at as high as around 600 passenger cars per 1,000 inhabitants, whereas late starters such as Japan or Spain reach saturation density levels between 250 to 300 passenger cars per 1,000 inhabitants.

Other Countries

The difference in the high saturation density of the "automobile bandwagon" (i.e., OECD) countries compared to the estimated world total implies that most of the developing countries (with most of the world's population) will not reach car diffusion levels anywhere close to that of developed countries. In order to corroborate this hypothesis we present below diffusion analyses for selected COMECON (Council for Mutual Economic Assistance) and developing countries to investigate whether a forthcoming saturation is also likely in these countries. With the support of empirical data, Tables 3.3.3 and 3.3.4 show that this is indeed the case, implying that car diffusion in these countries will not "tunnel through" the forthcoming saturation in industrialized countries. This also makes our scenario of global car saturation at very divergent density levels internally consistent.

The diffusion of registered passenger cars in centrally planned economies and developing countries is also analyzed using both absolute and relative measures. The diffusion trajectories from both measures conform well with each other. It is also noticeable that the ultimate diffusion densities which may be achieved will be as heterogeneous, or even more divergent, as those emerging from our analysis of industrialized countries. The uncertainty of the estimated saturation level is generally higher in these countries than for industrialized countries (see in particular the much wider uncertainty bands for the estimated saturation level K). In most cases, this is because the diffusion process is not yet close enough to the saturation level to allow for higher statistical certainty. In some of the additional countries analyzed (and not reported in Tables 3.3.3 and 3.3.4), in particular, India, Indonesia, South Korea, the Philippines, and Thailand, we were not able to determine an estimate for the saturation level, as the growth process is still in its early (exponential) phase. This implies that the ultimate saturation level for these countries might be larger than the present car density by at least a factor of three. However, considering the present low density levels of passenger cars per 1,000 inhabitants in these developing countries, e.g., 2 for India, 11 for Thailand and 16 for South Korea, even growth by a factor of three would not noticeably affect our average estimated saturation density of less than 70 cars per 1,000 inhabitants at the global level.

Of the developing countries analyzed, only Venezuela appears to have a motorization pattern that allows for significant growth well into the next century. Venezuela seems to be an example of a country that started early and grows slowly to high density levels. Taiwan, on the other hand, is similar to Japan: both countries were late-starters but caught up quickly as reflected by their rapid Δt of diffusion.

Table 3.3.3. Passenger car diffusion1) from 1950 to 1987 in COMECON and developing countries (in million passenger cars registered).

			Passenger Cars (millions)				
Country	t_0	Δt	K	K-range2)	Present (1986–87)	Present3) in % of K	R^2 of estimate
Czechoslovakia	1977	26.2	3.5	3.1–4.0	2.7	68–76	0.9942
GDR	1976	28.6	4.1	3.6–4.7	3.3	71–80	0.9995
Hungary	1979	21.3	1.9	1.6–2.1	1.5	73–82	0.9991
Poland	1985	25.6	7.2	5.1–9.2	3.3	36–46	0.9980
Argentina	1979	39.1	6.0	4.8–7.3	3.9	53–65	0.9970
Brazil	1979	27.8	13.6	12.6–15.8	10.1	64–74	0.9986
Mexico	1984	33.1	9.9	6.2–13.0	5.2	40–53	0.9768
Taiwan	1988	20.4	2.8	0.3–5.9	1.0	17–36	0.9985
Venezuela	1996	44.1	9.1	2.5–35.4	2.4	7–26	0.9994

1) Data period used for parameter estimation: 1950–1986/1987.
2) With 90 percent probability.
3) Range corresponds to high estimate and best fit case respectively.
(Data sources: see Table 3.3.1.)

Table 3.3.4. Passenger car diffusion1) from 1950 to 1987 in COMECON and developing countries (in passenger cars registered per 1,000 inhabitants).

			Cars per 1,000 inhabitants				
Country	t_0	Δt	K	K-range2)	Present (1986–87)	Present3) in % of K	R^2 of estimate
Czechoslovakia	1977	27.0	225.9	196.1–420.0	173.5	41–77	0.9936
GDR	1976	28.6	249.2	216.0–282.3	198.0	70–80	0.9994
Hungary	1979	21.9	178.3	154.7–202.0	144.9	72–81	0.9990
Poland	1985	27.3	194.7	136.2–253.1	112.1	44–58	0.9972
Argentina	1975	42.2	173.3	149.0–197.6	127.0	64–73	0.9894
Brazil	1977	32.4	96.7	83.6–119.5	74.0	62–77	0.9949
Mexico	1975	34.5	87.7	76.3–99.0	65.0	66–74	0.9593
Taiwan	1986	20.3	112.0	63.0–150.0	54.0	36–48	0.9981
Venezuela	1987	58.7	286.2	57.2–420.0	136.0	32–48	0.9919

1) Data period used for parameter estimation: 1950–1986/1987.
2) With 90 percent probability.
3) Range corresponds to high estimate and best fit case respectively.
(Data source: see Table 3.3.1.)

Growth of passenger car registrations will thus be very heterogeneous within developing and centrally planned economies. Different initial conditions, degrees of economic development, and the development of road infrastructures, explain the present diverse car ownership rates, even for similar per capita income levels. Countries like Argentina, Brazil, Mexico or South Korea have per capita GNP values of a similar order of magnitude (between 2,000 to 2,500 1982 US \$/capita in the Latin American Countries and 1,530 1982 US \$/capita in South Korea) whereas their car density levels differ by as much as between 64 to 124 cars/1000 inhabitants in the Latin American countries and 16 cars/1000 in South Korea (1986 values). From our point of view, diversity in diffusion levels will continue to be a characteristic feature of developing countries. The high growth rates in the urban population in developing countries calls more for the construction of efficient mass transit systems for short range travel. For long-distance travel, growth of air transport could be more consistent with the likely future developments in these countries than a linear extrapolation of past transportation (i.e., private car motorization) trends of industrialized countries.

This brings us to the last level of our analysis: the empirical test for two characteristic observations in the pattern of car diffusion. We have observed *acceleration* of diffusion speed and *decrease* in ultimate diffusion levels as a function of the "learning time", i.e., the time period between the beginning of the diffusion process and the time available for growth (i.e., Δt). Figure 3.3.9 summarizes the results of Tables 3.3.2 and 3.3.4 by analyzing the relationship between the length of the growth process of automobile diffusion (Δt) with the ultimate saturation level and the time a country began motorization. The acceleration of diffusion for late starters is not unique, since we observed earlier a similar relationship in the growth of the railway networks.

Perhaps the most striking finding that emerges from Figure 3.3.9 is the straightforward mathematical expression for the "acceleration" phenomenon of diffusion in relating the log of Δts to the time period in which private cars achieved a level of one percent of their ultimate saturation density. This functional relationship was first hypothesized by the late Ed Schmidt, 1983. *Schmidt's Law* describes the acceleration of the diffusion rates of latecomers and their lower diffusion levels. Using linear regression, we arrive at the functional expression shown in Figure 3.3.9. The data confirm the acceleration tendency at a statistically highly significant level, and explain as much as 89 percent (R^2 adjusted for degrees of freedom) of the variance in the observed data. If one extrapolates the catch-up tendencies, one arrives at the conclusion that a country starting the diffusion of private cars now, would have a diffusion rate Δt lower than 10 years. Such rapid diffusion of car ownership appears quite infeasible from a practical viewpoint, and could not in any case result in a significant growth in both absolute and relative car registration figures.

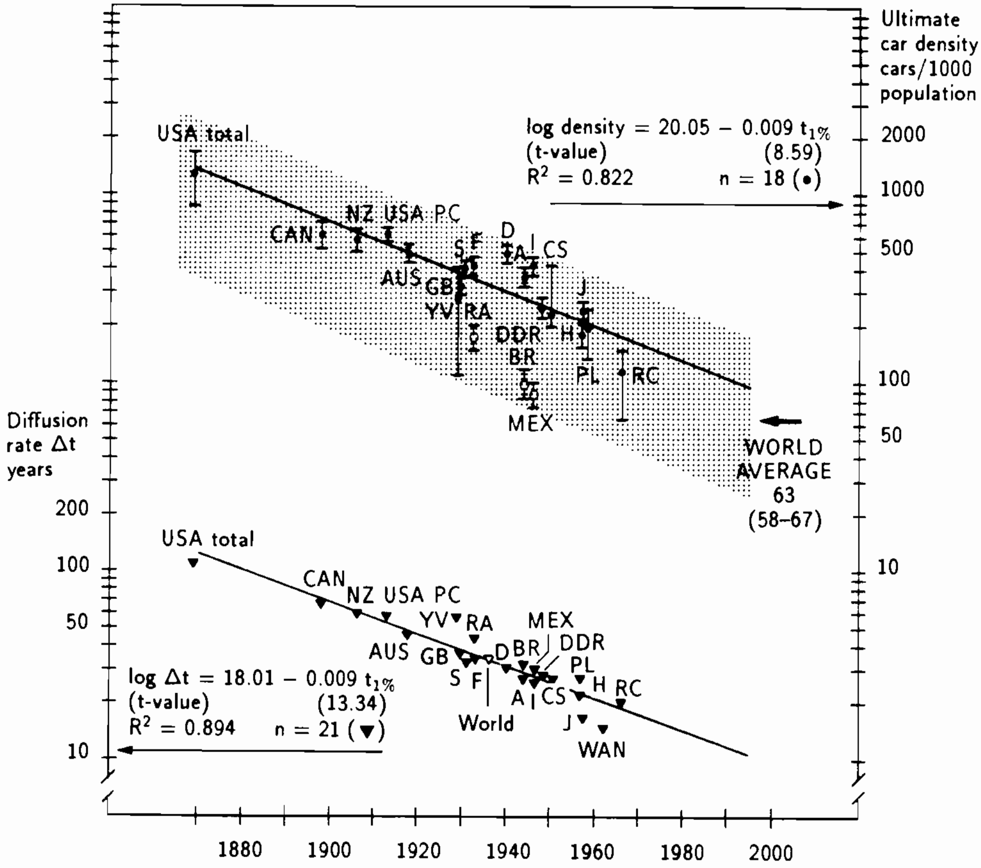


Figure 3.3.9. *Schmidt's Law*: Diffusion (growth) rates and densities of passenger car ownership as a function of the introduction date of the automobile.

We proceed to analyze the relationship between the estimated ultimate car diffusion level and the start of motorization. Early starters such as the USA* and Canada will have significantly higher car density saturation levels (above 500 passenger cars per 1,000 inhabitants) than countries like Hungary, Japan, or Taiwan, where density levels of about half the level appear typical. The estimated regression equation of the declining density trend explains as much as 82 percent

---

* Figure 3.3.9 also includes the diffusion of total road vehicles in the USA (i.e., riding horses and mules and later automobiles), because of the high density of individual road transport means in the US prior to the advent of the automobile. Recall here also that at the moment when the replacement of horses became complete, i.e., by the 1930s, passenger car density in the USA was already nearly as high as in present day Japan. Recall also the significant difference in the saturation levels of passenger car and total road vehicles density in the USA.

of the variance in the estimated saturation density levels of the industrial countries forming the "automobile bandwagon".

Developing countries can be regrouped into three categories. Countries like Argentina, Brazil, and Mexico appear to follow the same declining density trend as industrialized countries, albeit at a significantly lower level. Venezuela and Taiwan are midway between developing and industrialized countries. All the developing countries, however, seem to follow a declining density trajectory similar to industrial countries. Other developing countries like China or India fall even further below the density envelope of developing countries shown in Figure 3.3.9. However, in the same way that the USA does not provide a model for car diffusion in countries like the GDR or Japan, car diffusion in Brazil or Mexico does not imply a realistic model for Nigeria, India or even China. Density levels below 10 passenger cars per 1,000 inhabitants, which are consistent with our projected world average density figure, appear more likely for these countries.

The declining density envelope and the trend line of Figure 3.3.9 suggest that any country starting diffusion now with a rapid growth in car registration would not achieve a great deal with respect to the resulting diffusion levels. If we extrapolate the density envelope trends for the future, we arrive at the conclusion that an industrial country starting motorization now would grow very fast (Δt of around 10 years) but at the same time achieve a density level below 100 passenger cars per 1,000 inhabitants. Newly Industrialized Countries (NICs) on the other hand, may achieve with a Δt of 10 years figures close to our estimated world average of around 70 cars per 1,000 inhabitants, whereas "developed" developing countries like Brazil or Mexico might achieve density levels in the vicinity of 20 to 30 passenger cars per 1,000 inhabitants should motorization start now. For those developing countries that do not have the necessary initial conditions for motorization to take-off (primarily in terms of disposable income) like India or China, even these values are probably one order of magnitude too high: These countries fall outside the diffusion density envelope shown in Figure 3.3.9. We believe that within the next twenty years no trend in car densities, similar to those in industrial or newly industrializing countries, will emerge in developing countries such as China.

We postulate that the acceleration tendency in private car diffusion has therefore a number of implications. First, it appears unlikely that a similar diffusion will occur in countries that are not part of the present diffusion "bandwagon" of industrialized and industrializing countries. The second implication is that, with very few exceptions, the expansion in passenger cars registered will approach saturation by the turn of the millennium. A *season of saturations*, which is closer for late starters like Japan and Western Europe than for North America, appears forthcoming. This can be either positive or negative depending if we are looking at it from the point of view of the automobile industry or the environmental movement respectively.

The relative diffusion levels that ultimately result from motorization differ from country to country as a result of their different geographical, economic, etc., boundary conditions, and the related accumulated "experience" involved in the diffusion process. This appears all the more plausible if we consider that in countries where diffusion spans many decades, lifestyles and settlement and spatial organization patterns are developed that go along with high motorization levels (e.g., North America). In countries like Japan, where motorization occurred within 20 years similar cross-enhancing effects are much more limited and ultimate motorization levels are therefore considerably lower. Our observation on the different car densities in various countries is consistent with the results of our analysis on the spread of the railways as a dominant transport mode prior to the automobile.

The "clustering" of the saturation dates that emerges from our analysis, is all the more noteworthy in view of the diffusion times involved spanning several decades. This corroborates our working hypothesis of the "bandwagon" effect in the diffusion of a technological paradigm (as represented by the automobile industry) which reaches saturation within a relatively short time period. Based on such a conjecture one would expect a significant structural discontinuity in the evolution of the world automotive industry, a transition phase which could well be accompanied by a period of high market volatility and intensified international competition for survival in saturating markets. The acceleration of diffusion towards the end of this period of clustering saturations (i.e., the *Kondratieff barrier*) can also be observed in the diffusion of new technologies within the automobile industry itself, probably pointing at a deeper functional symmetry between automobile diffusion *per se* and technological change in the car industry itself.

Technological change in the car fleet

As in Sections 3.1 and 3.2, we conclude our discussion on the diffusion of roads and the automobile by presenting illustrative examples of technological change in the car fleet. A detailed discussion and analysis of innovations and technological change in the automobile sector of the USA is given by Abernathy, 1978; Blackman, 1974; Jutila and Jutila, 1986; and Nakicenovic, 1986, and we will not try to add examples to the comprehensive coverage provided by these authors.

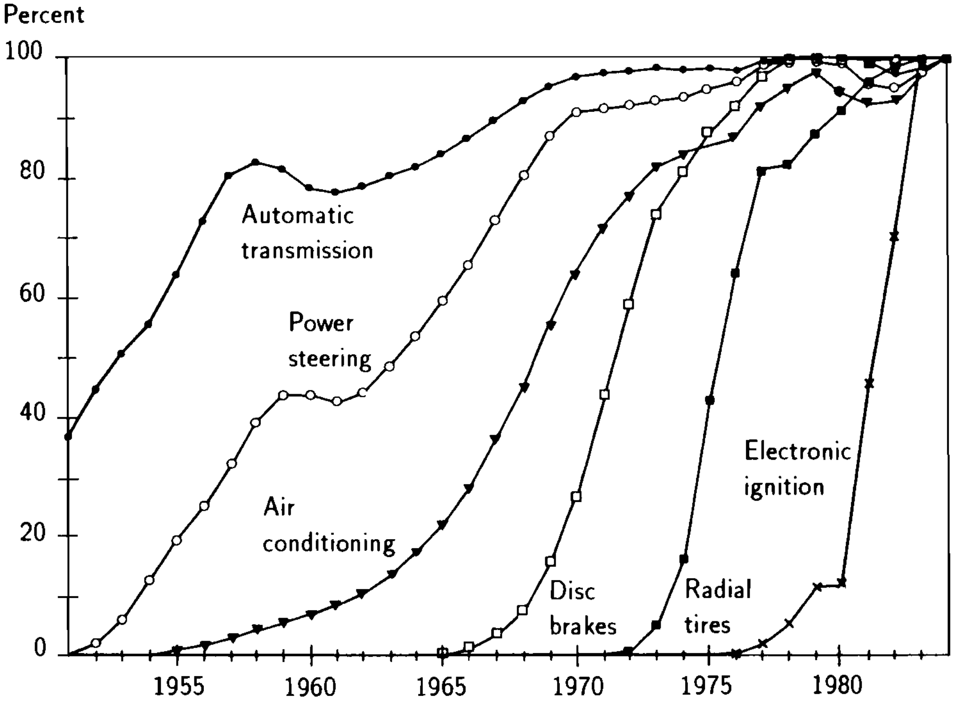


Figure 3.3.10. Diffusion of new technologies in the US car industry (in percent of car output). (Source: Jutila and Jutila, 1986.)

As illustrative cases we will consider the introduction of different technological innovations in the US car manufacturing industry as well as the introduction of compulsory emission control measures in the automobile fleet of the USA.

Figure 3.3.10 presents the diffusion of different technological innovations (automatic transmission, power steering, air conditioning, disc brakes, radial tires, and electronic ignition) in the US car industry. Whereas the functional pattern in the form of S-shaped diffusion curves is hardly surprising, we note an interesting acceleration tendency in the introduction of technological innovations when approaching the *Kondratieff barrier* in the saturation of automobile expansion for most countries identified above. It is as if the industry itself anticipates the approaching saturating markets and tries to gain ground by accelerating the pace of incremental innovations. Note that during periods of steady, regular high growth rates (the steep part of the automobile diffusion curve) the rates of technological change appeared to be much slower. Faced with almost saturated markets, the industry appears to be forced to rapidly introduce technological innovations as part of the competitive "elbowing" in the struggle for market shares. Thus there is certain empirical evidence of the innovation

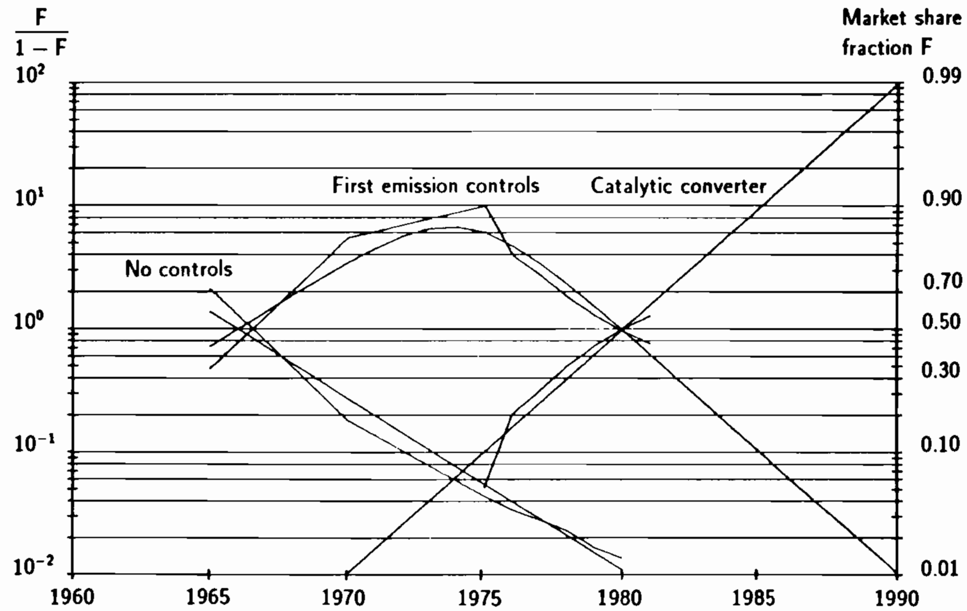


Figure 3.3.11. Substitution of cars with emission controls in the USA. (Source: Nakicenovic, 1986.)

triggering effects when approaching a period of market volatility and increased competition (i.e., the *season of saturations*).

As regards the automobile industry, a similar innovation "pull" can be observed in the early phase of development, which resulted in the drastic decline of nominal and real term prices of cars up to 1920 (see Chapter 4). This earlier period of rapid technological change of the, at that time still very young, automobile industry occurred during the *season of saturations* of the railway/steam/coal technological cluster described in Section 3.2 above. The very rapid replacement of horses by the automobile was without doubt made possible by the rapid technological improvements in cars, their increasing performance and comfort, and the decrease in prices in the the first 20 years of the automobile industry.

A final example of technological change in the automobile industry is presented in Figure 3.3.11, which analyzes the introduction of environmental control technologies in the US automobile fleet. The diffusion of environmental control techniques in the car fleet of the USA is interesting for several reasons. First, it provides a test as to whether the standard models of technological diffusion and substitution are also applicable for cases where comparative advantage does not (expressed in producers or consumers preference differentials) drive the substitution process, but rather, government regulations. In such cases, the substitution model does not describe the diffusion of a new technology in terms of how it becomes incorporated into production or consumption (for instance since 1973 all new cars were required to comply to low nitrogen emission standards, thus the percentage of cars produced/sold with the new technology, the catalytic converter, changed almost instantaneously from almost zero to 100 percent) but rather how it diffuses into the existing *stock* of a technological population. Figure 3.3.11 provides convincing evidence that technological dynamics resulting from (legal) regulatory measures can also be captured by technological diffusion and substitution models. In this case the models describe the *population dynamics* inherent in the car fleet of a country. As such it provides a simple mathematical model reflecting the change in a technological population stemming from the dynamics of obsolescence* and the resulting replacement rates, and that of market expansion.

The share of cars equipped with different environmental control techniques and analyzed in Figure 3.3.11 are: (partial) emission controls in the form of crankcase (hydrocarbon emission reduction), exhaust controls (for the reduction of carbon monoxide and hydrocarbon emissions), and fuel evaporation controls; catalytic converters; or no controls at all. The relative share of cars with any one of the three types of emission control technologies in the automobile fleet of the USA can thus be described by a multiple logistic substitution model. The Δt of the diffusion of cars equipped with catalytic converters is about 10 years, indicating that the average life time (although varying over time) of cars in the USA can be approximated by the estimated Δt. The analysis also indicates that there are good reasons to be optimistic about the time horizon of the positive effects of introducing similar environmental control technologies in other countries, e.g., catalysators recently became mandatory in Austria. It is certainly not incidental** that the time constant involved (around 10 years) in solving a serious environmental problem resulting from existing road

---

* Marchetti, 1983, provides a (logistic) "death curve" of the survivors of an age cohort (model year 1967) in the US car fleet. The Δt of the logistic mortality curve is 9.2 years, indicating that it took 9.2 years before the number of the 1967 models originally sold decreased from 99 to 50 percent or from 90 to 10 percent. The Δt is thus an expression for the average survival time of cars of a given cohort (model year).

** The most straightforward explanation for the similarity in the dynamics of the introduction of catalytic converter cars and the replacement of horses in the transport sector (Δt of around 12 years) would suggest similar useful lifespans for the two technologies. The average life time of horses (and carriages) was in the order of around 12 years (the average active life expectancy of a horse being between 12 to 15 years, Meißel, 1988) which is in fact not so different from the current average lifetime of a car. This is a further indication of a deeper underlying functional similarity between the automobile and the horse in road transport.

transportation is similar to the dynamics of resolving earlier environmental problems resulting from road transport (horse manure), as presented in the analysis of the substitution of horses by automobiles in the USA at the beginning of this century (Δt of 12 years).

Summary

Figure 3.3.12 summarizes the diffusion envelopes resulting from our analysis of the two phases of automobile diffusion in industrialized countries. The horse replacement and car diffusion trajectories are summarized by plotting the respective logit transformation of their diffusion and substitution curves. Together these diffusion curves represent international diffusion bandwagons, i.e., rather narrow bands, in which all the diffusion processes of the industrialized countries analyzed are confined. Figure 3.3.12 identifies two such diffusion bandwagons. First, the "horse-replacement bandwagon", in which the diffusion rates are swift and very uniform in all countries and which reaches saturation in the 1930s.

The second one, consisting of the diffusion of passenger cars, is progressively converging towards the saturation period, showing a distinct acceleration for those countries that started motorization later. As such, Figure 3.3.12 provides a summary of the diffusion of the technological "paradigm" associated with the "automobile or petroleum" age. Figure 3.3.12 also incorporates the diffusion trajectories of other infrastructures and technologies representative of the "automobile and oil age". These include the growth of surfaced road infrastructure in the USA and the USSR and the diffusion of the transport infrastructure for petroleum and oil transport (oil pipelines) in the USA (Grübler and Nakicenovic, 1987); another example returns to older transport technologies (ships) by showing how the internal combustion engine substituted coal based steam propulsion technology in the UK. The associated large infrastructures of the "oil age" have tunneled through the first saturation phase of the growth of the automobile. In turn, these infrastructures appear to be approaching saturation around the year 2000, which is in tune with the cluster of saturations for the growth of passenger car registrations in all industrialized countries.

Figure 3.3.12 thus presents a condensed summary of the growth of a technological paradigm associated with the use of oil and the internal combustion engine as prime mover, either in creating its own market (automobile diffusion) or replacing existing techniques like horse carriages, or coal fired steam engines in the ship fleet or in the railway sector. Although the picture appears rather complex, one can distinguish three classes of diffusion "clusters".

- Very long term, represented by the growth of infrastructures (oil pipelines and surfaced roads in the USA and the USSR), the replacement of steam by motor ships (UK), and the diffusion of the automobile in North America.
- An accelerating diffusion cluster represented by the automobile growth in most industrialized countries (Austria, Australia, France, Germany, later the FRG, Italy, Japan, Sweden, Spain, the UK, and the second growth pulse for passenger cars worldwide).
- Short (rapid) diffusion processes represented by the replacement of horses by cars in France, the UK, and the USA as well as worldwide (first growth pulse of cars) and the diffusion of catalytic converter cars in the USA (and returning to railways, the replacement of steam locomotives by diesel and electric locomotives).

Despite overlap in these diffusion processes and although their different Δts result in a complex pattern of rates of technological change, one can clearly identify two main periods of structural discontinuity characterized by a clustering of the saturation phases of various diffusion processes. These *seasons of saturations* can be identified in the period 1930 to 1940 (saturation of the replacement process of horses as a means of road transport) and in the period from the mid-1980s to shortly after the year 2000, which is characterized by the progressive saturation of the car/oil technological cluster. In fact, only few diffusion processes appear to "tunnel through" this saturation "barrier" which is most notably* reflected in the growth of surfaced road infrastructure in the USSR. These again would start to enter their proper *season of saturations* some 50 years after the beginning of the *saturation season* in the 1980s.

Our argument with respect to discontinuities in the rate of technological change and diffusion and their respective impacts on economic growth, does not rely on a rigid synchronization or "focusing" of either the introduction, growth or saturation periods of *all* processes. It is sufficient that a number of important replacements and diffusion processes take place at the same time in a number of core countries, leading to a prolonged period of economic growth. The growth stimulating effects become progressively exhausted as the

---

* Automobile diffusion in Canada, the USA and New Zealand may continue for some time even after the saturation of motorization in industrialized countries, although with decreasing growth rates.

replacement and diffusion processes start to saturate during relatively short time periods, resulting in periods of retarded growth, recession, economic restructuring, and later on in the introduction of new technological and social innovations. In Chapter 5 we will develop a new measure for long-term discontinuities in the technical, economic, and social change of a country in aggregating a large number of growth and transformation processes. We will show that indeed strong discontinuities emerge, even with only a relatively weak "clustering" in the introduction and saturation phases of many processes of change.

The exhaustion of the growth potential of the oil/automobile "cluster" will lead to saturating markets and intensified competition. However, it is more important to realize the increasing disbenefits associated with the further intensification of an "automobile society". This includes in particular the negative environmental externalities associated with dense car populations in urban agglomerations. At the same time automobiles continue to be a symbol of individual mobility and wealth, and as such the further diffusion of the automobile may ultimately be blocked by the emergence of new social values and attitudes.

The start of saturation (i.e., the growth to limits) of the largest part of the technological cluster associated with the internal combustion engine and oil usage, that emerges from Figure 3.3.12, is probably best characterized by the fact that the market share of oil in global primary energy use is saturating at present (see Figure 2.1.3 above). From such a perspective, a major structural discontinuity in the evolution of a technological paradigm, responsible for much of the economic upswing after World War II, appears consistent. Therefore, the resulting (energy) market and price volatility in the 1970s and 1980s should not be too surprising.

Yet, it is exactly during these periods of discontinuity and volatility that a new technological paradigm is being shaped. In the next Section we will discuss air transport and the fact that there are some indications that air transport may be part of a forthcoming sociotechnical paradigm "tunneling through" the forthcoming saturation phase of the automobile.

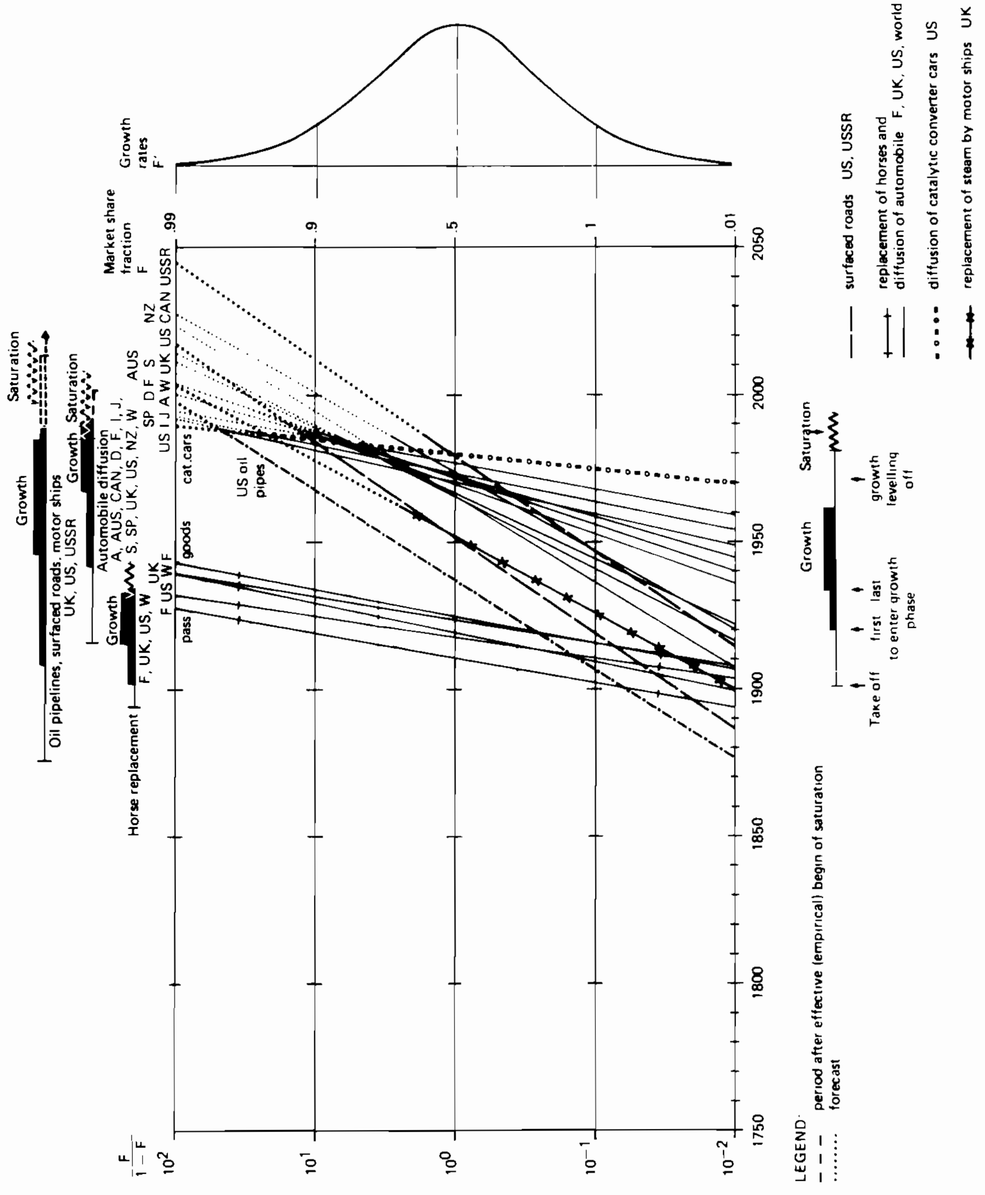


Figure 3.3.12. Expansion of the oil/internal combustion engine technological cluster: diffusion of roads, automobiles, oil pipelines and motor ships in core countries. *A* – Austria, *AUS* – Australia, *CAN* – Canada, *D* – FRG, *F* – France, *I* – Italy, *J* – Japan, *NZ* – New Zealand, *S* – Sweden, *SP* – Spain, *UK* – United Kingdom, *USSR* – USSR, *US* – USA, *W* – World. [Source: Oil pipelines (Grübler and Nakicenovic, 1987), motor ships in merchant ship tonnage of UK (Nakicenovic, 1984), all others (authors' own estimates).]

3.4. Air Transport

Air transport is the most successful (if measured in terms of market volume growth) existing long-distance transport mode. Commercial aviation dates back to the 1930s when the first airplanes achieved the productivity, range, and economic capabilities necessary to compete with the more traditional long-distance transport modes – railroads and ocean-liners. Ever since, air transport has eliminated these older technologies in almost all market segments of passenger transport for long distances (above 1,000 km). This is not only true for developed countries but also for many areas of the third world.

Once a new transport mode makes possible the extension of the spatio-temporal range of human activities, a very strong impact on travel demand can be expected. The bottom of Figure 3.4.1 illustrates such a technological revolution in the transport sector: The replacement of transatlantic passenger crossings by sea first by the piston and later the jet aircraft. As can be seen from the Figure, it took around 30 years to completely replace ships by aircraft. The top of Figure 3.4.1 shows the same process of technological substitution in analyzing the total number of passengers transported by the three technologies. The peak of transatlantic passenger crossings by ship occurred in the late 1920s, with over 2 million passengers transported annually (Woytinsky, 1927). The emergence of the piston propelled aircraft resulted in a 50 percent decrease in the number of passengers previously crossing by ship, whereas the total market volume did not change noticeably. Thus, by the mid-1950s when the jet aircraft was introduced, around 1 million people were crossing the Atlantic by ship and an equal amount by piston aircraft. Much more decisive, however, was the impact of the jet aircraft on the number of passenger crossings. Market volume increased dramatically in response to the availability of a new, fast, and convenient form of transatlantic travel made possible by the jet aircraft. Compared to its pre-introduction time, the jet aircraft increased the volume of transatlantic crossings by a factor of 10 over a period of 30 years.

The rapid expansion of air transport was only made possible by a large number of scientific and engineering achievements in aircraft technology, including aerodynamics, aircraft design, propulsion technology, the development of the necessary air transport infrastructure (air control and air routes), communication (radio and radar) and "transfer" infrastructures (airports and hubs), linking air transport to other transport modes. Another decisive factor which enabled the rapid growth of air transport was the great improvement in the productivity (size and speed) of aircrafts.

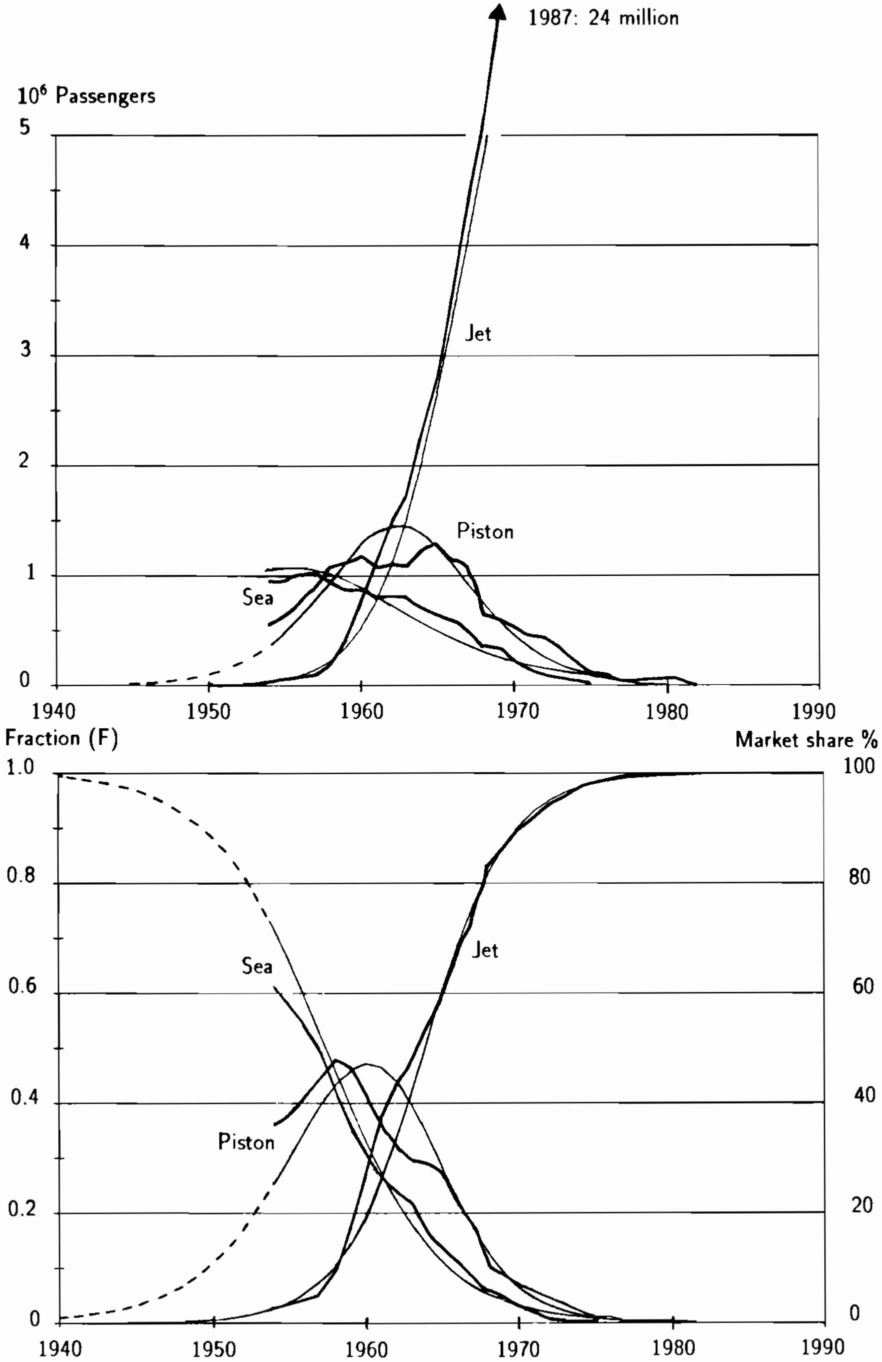


Figure 3.4.1. Fractional shares of ocean ships, piston, and jet aircrafts in transatlantic passenger crossings (bottom); number of passengers crossing the Atlantic by the three transport modes (top); (empirical data and estimates based on logistic substitution model). (Data source: IATA, 1972, and 1987.)

The engineering developments in aircraft design* are probably best symbolized by the DC-3 airliner, the first "modern" passenger aircraft introduced in 1935. The widespread use of this aircraft (DC-3s accounted for about 20 percent of the aircraft fleet of major US air carriers in 1936, by 1945 this share had risen to over 90 percent, as shown by Martino *et al.*, 1978) marks the beginning of the "take-off" of large-scale commercial aviation. The success of the DC-3 was not only the result of superior engineering and technological performance but was also due to the fact that it corresponded much better to the requirements of a rapidly expanding market. The DC-3 aircraft could carry about twice as many passengers as other aircraft models it replaced (e.g., the Ford Trimotor or the Boeing B-246). Other technological innovation milestones that were important for the growth in the performance of aircraft, include improvements in the performance of piston engines (discussed below), an elaborate communication infrastructure (in particular the use of radio), and subsequently, the availability of radar.**

The aircraft industry and air transport in general can be considered as the first truly global transport technology and infrastructure. In fact, the number of manufacturers competing in the world aircraft market is rather small, and even more limited in the area of aircraft propulsion, where, with the exception of the USSR, only three major jet engine manufacturers exist today. Successful aircrafts like the DC-3 or the B-747 (Jumbo) were, and are used in even the most remote areas of the globe. As is appropriate for the global nature of air transport, we discuss here the evolution of air traffic and of technological change in aircrafts primarily at the global level.

World air traffic volume

Figure 3.4.2 presents the growth of the world air transport volume measured in t-km transported annually. The figures include centrally planned economies and all commercial carrier operations (scheduled and others), but exclude (as no international statistics are available)

---

* The first full metal aircraft suitable for longer distance passenger transport was the German Junkers F 13 (Lufthansa, 1986).

** Before the invention of radar, the main obstacle for the early air transport industry was that flights could take place only during day time. This technological barrier to further market growth was recognized in a by now classic study under the leadership of the sociologist William Ogburn in 1937 (National Resources Committee, 1937). In this study a large number of potential technological "fixes" to overcome this problem were identified, without however anticipating the final solution: radar, which was invented in the 1940s. Without taking a too optimistic viewpoint about the feasibility of "technological fixes", this case illustrates nevertheless a "market pull" impact on innovative activities, which finally helped to overcome a barrier to further market expansion.

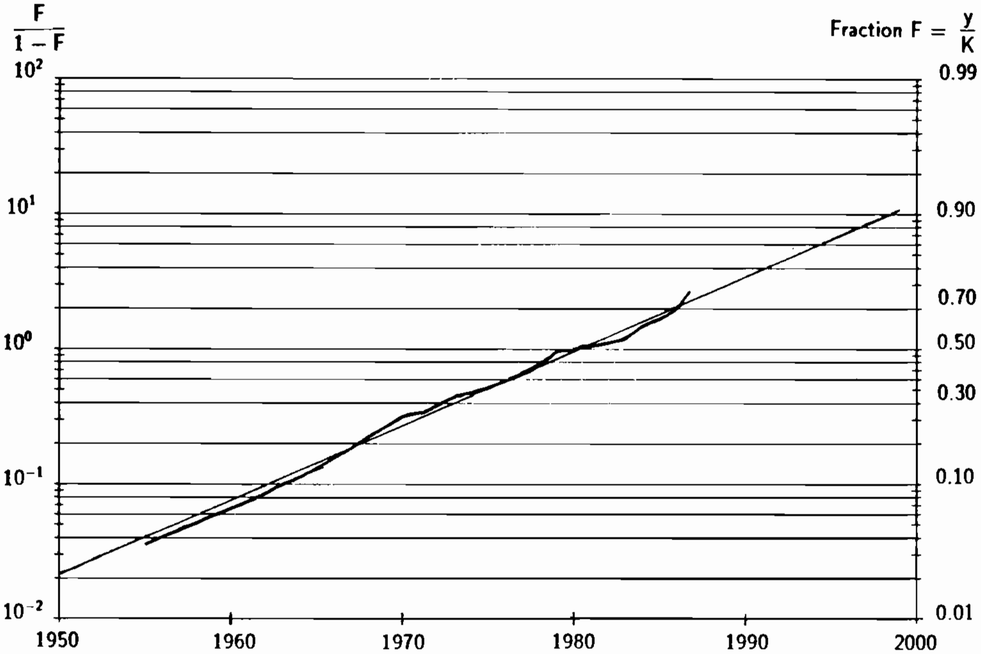


Figure 3.4.2. Volume of air transport worldwide (including centrally planned economies) passenger, cargo, and mail (in 10^9 t-km per year) [K 289.2 10^9 t-km; t_0 1980; Δt 34 years]. (Data source: IATA, 1961 to 1988, and own estimates.)

private commuter traffic. The total transport volume expressed in t-km includes cargo and air mail as well as passenger traffic. Passenger-km are converted to t-km based on a weight of 90 kg per passenger including baggage, following International Air Transport Association (IATA) conventions. Total ton-km transported worldwide* by air have increased by approximately two orders of magnitude to the present level of 217.3 billion (10^9) t-km. This market volume growth proceeded along the logistic trajectory estimated from the empirical data and plotted in Figure 3.4.2.

The growth of the t-km (including passengers) transported by all air carriers increased at an exponential rate up to 1980, with a doubling of the transport volume every few years. As of 1980 however, only one doubling appears to be left until the estimated saturation

---

* Data sources for global air traffic: IATA, 1961 to 1988, but covering also non-IATA member operations. The USSR has been included in these statistics since 1973. We have supplemented IATA figures with data on the USSR prior to 1973, using national Soviet statistics (Narod. Khoz., 1982, and 1986, and Lewytzkyj, 1979). As for the 1950s and early 1960s non-scheduled air carrier operations are excluded in the IATA statistics, we have estimated these operations by assuming that 20 percent of all operations (similar to the situation in the late 1960s and early 1970s) are accounted for by non-scheduled operations.

level of around 290* billion t-km will be reached after 2010. The dynamics of the growth process in total volume transported is in agreement with independent estimates made for passenger traffic alone (e.g., Nakicenovic, 1988, estimates the growth of passenger traffic to proceed with a Δt of 29 years and t_0 in 1977 to an ultimate saturation level of around 200 million passenger-km per hour). If this growth trend in world air traffic continues, it will have important consequences on the necessary support infrastructures; air traffic control, airways, airports, and hubs. Further market volume growth, which appears likely to continue its historical pattern well after the year 2000, with a possible second expansion phase starting thereafter, could therefore be confronted with serious bottlenecks stemming from already congested air corridors and airports in many areas of the world.

Within the general growth in total market volume, one can note a gradual shift in the three different types of payload transported by air. In the beginning, when comfortable (closed body and later pressurized cabins) airplanes were not yet developed, most airplanes transported the highest value goods of low weight, where speedy delivery was important: information, in the form of mail. In the 1920s for instance, around 90 percent of all airplane kilometers flown and tonnage transported worldwide** was for mail transport. Mail flights were particularly important in the USA where aircraft finally replaced the legendary "pony express" for urgent mail deliveries in the West. Passengers and cargo started to become important only in the 1930s, following the significant improvements in aircraft design and technology.

Figure 3.4.3 shows the share of passengers, cargo, and mail in the scheduled t-km transported worldwide (including the USSR since 1973), organized with the help of a multiple substitution model. The gradual decrease in importance of mail in the t-km transported and the long-term structural shift between passenger and cargo transport becomes visible. Based on the model forecast, one could expect that already by the year 2000 some 30 percent of worldwide t-km transported by aircraft will be cargo. The growing importance of air transport in the goods sector, usually ignored as marginal in the tonnage oriented national transport statistics, is more plausible if we consider further shifts (dematerialization) in the direction of higher value goods for advanced economies (discussed in more detail in Chapter 4). Increased transport speeds and flexibility will become even more

---

* 90 percent probability that K will be between 243 to 335 billion t-km. R^2 of estimate: 0.993.

** Woytinsky, 1927, estimates the number of kilometers flown by aircraft to be approximately 4 million km in 1920; this figure increased at a rate of 40 percent per year to over 15 million km by 1924.

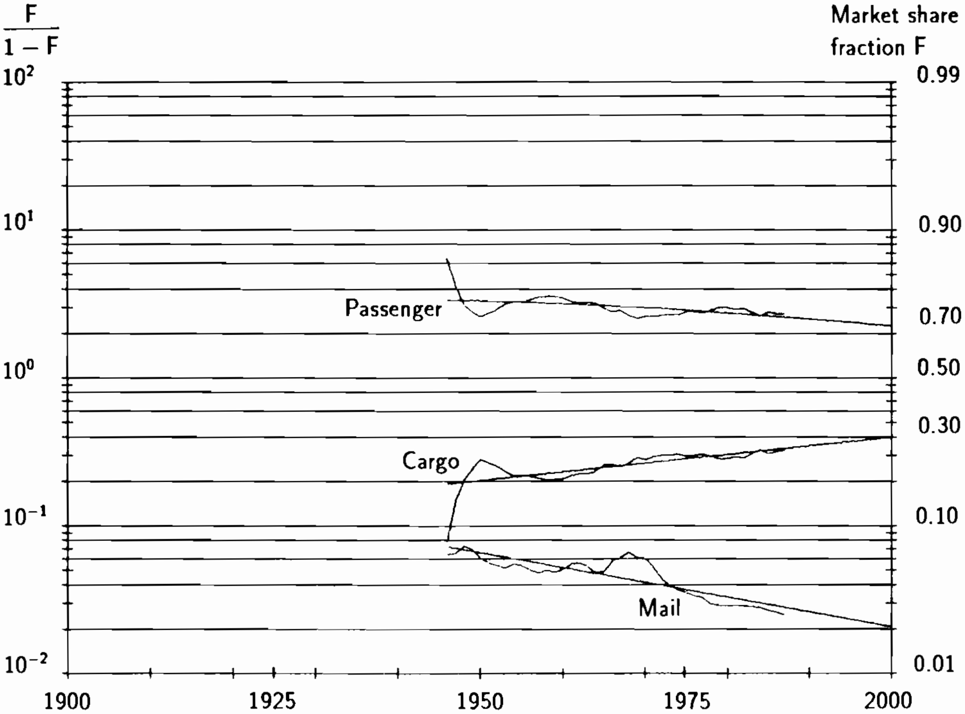


Figure 3.4.3. Share of passenger, cargo and mail in ton-km transported worldwide (scheduled services only, including the USSR since 1973). (Data source: IATA, 1972 to 1987.)

important under industrial production organization schemes following the "just in time" principle and the ever continuing quest to reduce (capital intensive) inventories. In addition, air transport may, in future, not be confined to traditional high-value or perishable goods like fashion garments, electronics, flowers, vegetables, etc., but may enter into markets considered hitherto as unimaginable for air transport. An excellent example is provided by Lufthansa (FRG), which (together with Alitalia, Italy) transports each week two Jumbos filled with car bodies (the Cadillac Allanté manufactured by Pininferina) from Torino, Italy to Detroit, USA for final assembly. One might consider this "air lift" of semi-finished car bodies as a production line extending over nearly 6,000 km. In a similar way Japanese car exporters to the USA also began to use air transport. The twice as high transport costs (compared to sea transport) being largely compensated for by lower insurance costs, reduced damage during transport (requiring previously expensive repair operations at the point of delivery), no requirements for protective coating, and finally, due to the higher transport speeds a considerable reduction of inventory (capital) costs.

The comparative advantage of air transport as regards speed, flexibility, and quality of service appears to be so high, that external events have not significantly affected the market volume expansion along the logistic trajectory presented in Figure 3.4.2 above. It is also worth noting that the regular growth pulse in world air traffic was apparently not influenced significantly by the rapidly rising fuel costs that resulted from the considerable appreciation of crude oil prices in the 1970s. This was most likely due to two factors. First, the demand for long-distance air travel is certainly amongst the least price sensitive transport market segments. Secondly, rising fuel costs appear to have been, to a large extent, compensated for by additional rationalization measures by the aircraft operators. It should also be remembered here that on average in the lifetime of an aircraft, about one third is spent on the aircraft itself, one third on replacing the jet engines (approximately every 10,000 operating hours) and the remaining third on operating costs (primarily kerosene). Although the USA is a special case (as regards competitive pressure in the air transport sector and resulting low air transport prices), no effects on the average transport costs per passenger-km of the "oil shocks" are noticeable. Real term air transport costs have been on a steady decline ever since the beginning of commercial aviation (see Chapter 4 below): from around 28 US cents per passenger-mile to below 5 US cents per passenger mile (in constant 1967 US dollars) between the mid 1920s and 1980s.

Air transportation is carried out through a spatial hierarchy. Short-distance domestic or short-distance international flights, as in Europe, usually provide the necessary feeder functions for air transportation over larger distances, i.e., on a continental or even transcontinental scale. We may therefore assume that whatever growth rate is realized at the higher spatial hierarchies of air transportation, will also result in an equivalent growth at lower hierarchical levels providing feeder services. To illustrate the complementary character of long- and medium-distances in air transportation, we show in Figure 3.4.4 the evolution of the international ton-km transported by air (passengers are again included on the basis of 90 kg per passenger including baggage).

Europe, with its many small countries and resulting short distances in international air transport, has been considered as functionally analogous to the domestic traffic in larger geographical spaces such as North America (i.e., as "domestic" air traffic in Figure 3.4.4). As can be seen, domestic (short- to medium-distance) flights have, after initial predominance over long-distance international flights, reached parity with international air transport volume since the early 1970s. We consider this as evidence of the complementary character

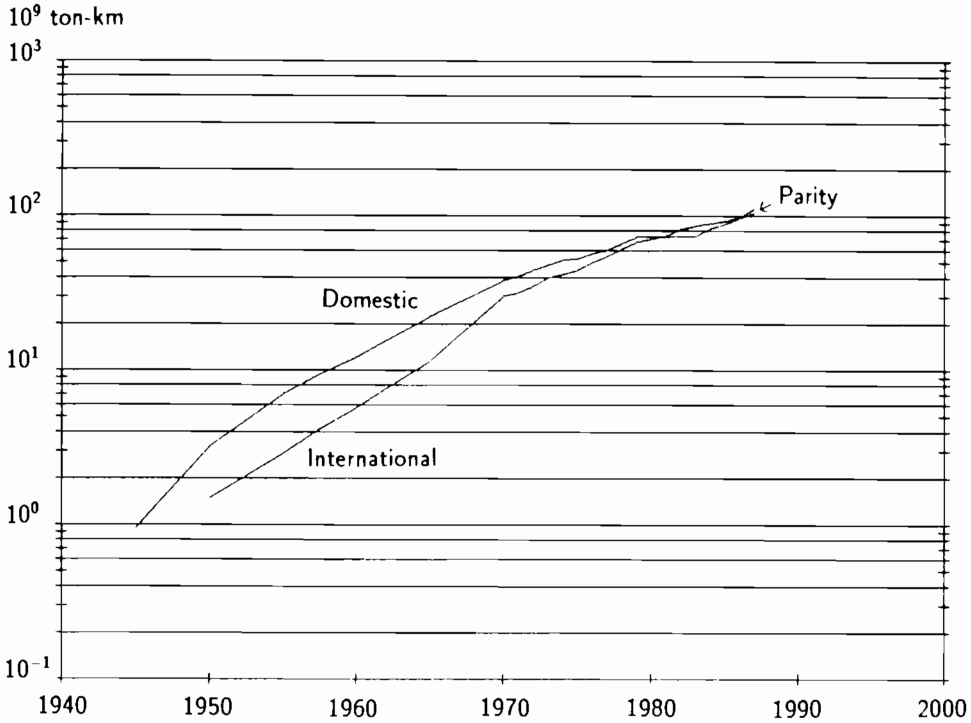


Figure 3.4.4. Growth of domestic and international air transport volume (passengers, cargo, and mail) including centrally planned economies (in 10^9 ton-km flown). (Data source: see Figure 3.4.2.)

of short- and medium-distances, and long-distance air traffic. Future growth of world air transportation may thus well be divided equally between international and domestic destinations. Functionally, Europe is considered to be a "domestic" geographical entity in such a scenario, which in view of 1992, does not appear to be an unreasonable working hypothesis.

As regards a regional breakdown in the volume of air transport, we face a serious data problem, in that no consistent data on the regional disaggregation of all air transport operations is available. For international, scheduled air traffic, IATA statistics* do allow, however, the analysis of its evolution by major geographical area, as shown in Figure 3.4.5 below. An interesting finding from such an analysis is the stability in the regional breakdown of international air transport operations of IATA members. Only the share of Europe declines from around 20 percent of the market share in 1949 to the present level of 12 percent. But as discussed above, European air

---

* Regional breakdown based on scheduled, international passenger-km flown. (Data source: IATA, 1961 to 1987.)

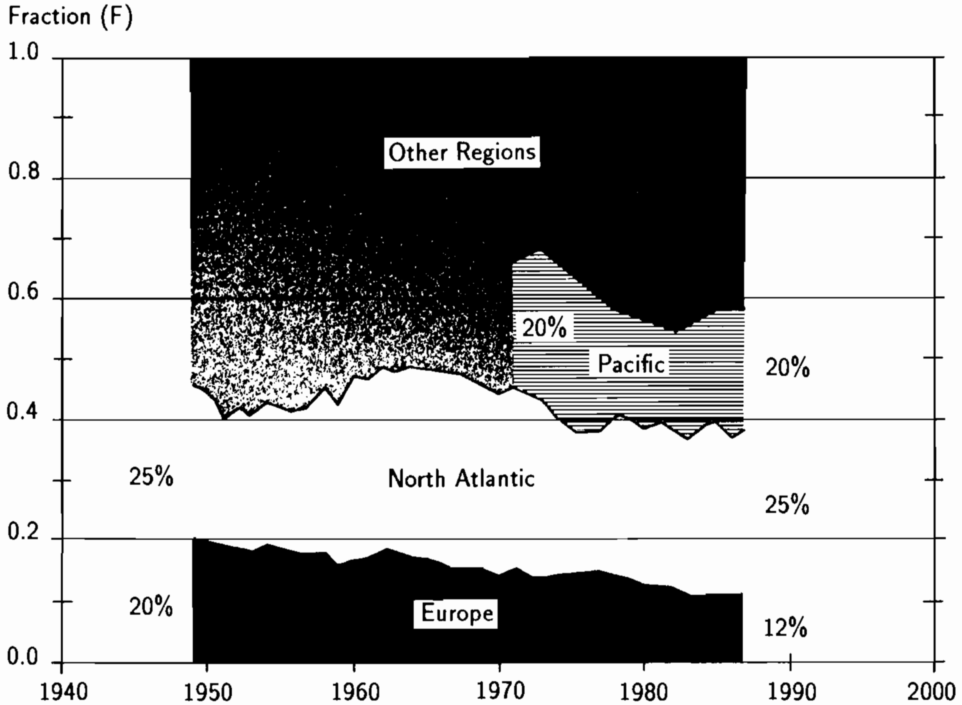


Figure 3.4.5. Share of different regions in scheduled international IATA air traffic.

traffic may be functionally better considered as a "domestic" air operation. Particularly noteworthy is the fact that despite the much discussed emergence of the Pacific rim countries, their share in international air transport volume has remained constant at a level of 20 percent ever since the first time statistics were collected for that region.

The growth of air transport described so far neither appears to be constrained by external events like rapidly rising fuel costs nor by the responses of the competitors challenged in long-distance travelling (ocean liners, cars, and railways). This points to the fact, that the comparative advantage of aircrafts (mainly their higher transport speeds) is so strong that it effectively compensates for any intervening external influences. The only remaining constraint to be analyzed is whether the development of aircraft *infrastructure* did*, or could in the future possibly constitute a bottleneck for the growth of air traffic. Bearing this in mind, we also ask whether technological development

---

* Contrary to today's situation of congested air corridors and airports, the land and air based infrastructure did not impose a significant constraint on the historical growth of air traffic volume.

and the improvement in *performance* of aircrafts kept abreast with the rapid growth of market volume.

Evolution of aircraft performance

The global fleet of commercial aircraft can be described in a number of ways. An obvious feature is the number of commercial aircraft in operation worldwide, as was described above for the growth in the number of passenger cars registered. The number of registered commercial aircraft (excluding private commuter planes) increased from about 3,000 in the 1950s to some 8,000* in the 1980s. At the same time, however, and contrary to the situation with cars, the combined performance in terms of size (carrying capacity) and speed of aircraft increased by about two orders of magnitude. The DC-3 as the first modern aircraft, could carry 21 passengers at a speed of around 350 km/h, i.e, an hourly *throughput* of 7,400 passenger-km. On the other hand, the Boeing 747 (Jumbo) introduced in 1969 can carry as much as 500 passengers at an average speed of 1,000 km/h, i.e., its performance is 500,000 passenger-km per hour. The largest planned B-747 (500 Series) will carry almost 700 passengers. The performance of passenger aircraft was therefore improved by a factor of 100 over the last 50 years. The size of the fleet is therefore not the most important indicator. One reason is because of the changing performance of aircrafts in service but also because much of the traffic is allocated to the most productive aircraft operating among the large hub airports, while other aircraft constitute the feeder and distribution system bound for destinations with lower traffic volume. The performance criterion of how fast and how many passengers an aircraft can carry, i.e., its productivity level in terms of passenger-km transported per hour, is analyzed in Figure 3.4.6 for long-distance aircrafts.

The performance of all commercial long-distance passenger aircrafts appears to evolve within a narrow "feasibility band" over time (measured by the year it was commercially introduced). It is interesting to note that all commercial aircrafts, in their productivity measure fall within this narrow "feasibility band", and that all commercially successful aircrafts fall exactly on the upper limit of the performance feasibility envelope depicted in Figure 3.4.6. This performance feasibility limit can be described by a logistic function (note that in Figure

---

* Available statistics are affected by uncertainties due to double counting of aircraft conversions and differences between aircraft in stock and effectively in use. Boeing, 1988, for instance reports worldwide some 8,640 aircraft effectively in stock, whereas MBB, 1989, estimates some 7,460 aircraft flying worldwide.

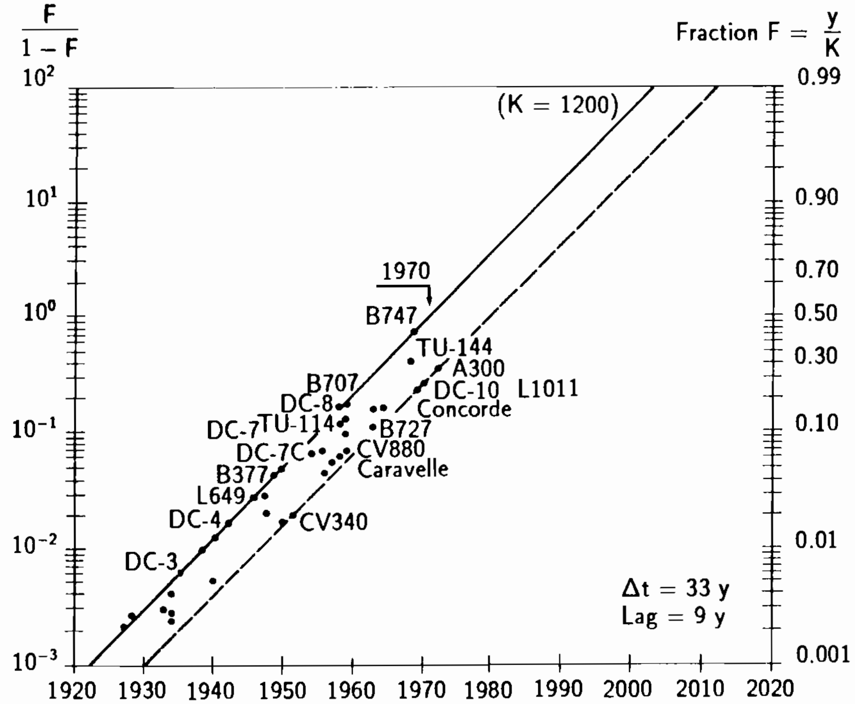


Figure 3.4.6. Passenger aircraft performance (measured in passenger-km per hour) [K 1200; t_0 1970; Δt 31 years]. (Source: Nakicenovic, 1988.)

3.4.6 the level below 1 percent of the ultimate saturation level is also represented, in order to capture the early phase of the development of commercial, long-distance passenger aircraft). The upper curve in Figure 3.4.6 represents a feasibility frontier of aircraft performance: all commercial aircraft introduced were either exactly on this performance limit or below. The aircraft designs, which were at their time of introduction, exactly on the "feasibility frontier" were all commercially successful. Most of the aircrafts falling below this feasibility limit proved (if not conceived for specific market niches, such as short-distance transport) commercially unsuccessful. This is best illustrated by the Concorde, which significantly fell short of the productivity of the B-747 introduced around the same time period. The Concorde turned out to be too small, and would in fact have required around 250 passenger seats (compared to the actual 100) in order to achieve the same productivity level as a B-747. From this perspective the commercial failure of the Concorde – despite being a technological marvel – may be better understood.

Thus, at any given time there appears to be only one appropriate (best) productivity level for long-distance commercial passenger aircrafts. Due to the fact that all commercial aircraft now travel at about the same subsonic speed (1,000 km/h) further improvements of productivity with existing propulsion technology will only be possible by increasing the size of aircrafts. According to the estimated curve in Figure 3.4.6, the asymptotic capacity for the largest aircraft would be about 1,200 10^3 passenger-km per hour (or about 10 billion passenger-km per year). Such large aircrafts may never be built, because of the logistics problems at airports for boarding over 1,000 passengers into an aircraft. This could be an indication of the fact that the next generation of long-distance aircrafts will develop in the direction of higher transport speeds (the second variable entering the productivity measure discussed here).

Another interesting observation from Figure 3.4.6 is that the productivity levels of all passenger aircrafts have evolved over time within a narrow "feasibility band". This is defined by the highest productivity aircraft (the feasibility frontier), trend line, and a second parallel curve (lagged by 9 years), which represents the *logistic growth of passenger-km* transported worldwide, expressed in the same measure as for aircraft productivity, i.e., passenger-km per hour. The conclusion from this observation is that aircraft productivity and market expansion evolved in unison and market growth did not appear constrained by a lack of availability of aircrafts at the appropriate productivity level in the various phases of the expansion of air passenger traffic. Another aspect of this interdependence of market and aircraft productivity growth is that no aircraft design exceeded the performance levels required by the expanding market at any particular point in time. The evolution of both aircraft technology and of air traffic volume is therefore characterized by a process of interrelatedness with no one-sided "market pull" or "technology push" prevailing, pointing at complex feedback mechanisms and interdependencies between market growth and the development of aircraft technology.

It took about 30 years for the performance of the most productive aircraft to increase from about one percent of the estimated asymptotic performance level to about half of that performance level. The DC-3 represents roughly the one percent achievement level and the B-747 roughly the 50 percent mark. Thus, in the beginning (i.e., in the exponential growth phase prior to the inflection point) the productivity of aircrafts was doubled every few years, whereas after the inflection point only one doubling in the performance level remains to be achieved. The implications of differentiating between these two periods in the development of aircraft performance on aircraft designs are straightforward. Prior to the inflection point (before the 1980s) aircraft design evolved in an "evolutionary" sequence of many models with increasing performance levels every few years. After the inflection point, when only one doubling of the performance level remains to be achieved (at least in the present phase of the development of aircrafts), a new "gradual" phase of aircraft development, consisting of modifications in existing designs, characterizes the industry. Thus, for the next two decades one should not expect a large number of new long-range aircraft models to appear on the market, but rather a gradual development, consisting of successive improvements and modifications of existing models. Because the B-747 can in principle be "stretched" by a factor of about 2, it could remain the largest long-range aircraft in the medium-term future (i.e., the next 20 years). This "stretching" policy is actually being followed in the development of the next generation of the B-747; which is planned to have around 700 passenger seats.

Let us make a final corollary from the observation of the parallel evolution of passenger market volume and the performance level of aircrafts. If both the market and the aircraft "throughput" in terms of passenger-km per hour evolve at the same rate, it means that the market volume can be transported by a constant number of aircrafts. Thus, instead of increasing the number of aircrafts in order to transport higher "fluxes" of passenger-km per hour, the productivity of a (near constant) long-range aircraft fleet grows accordingly. A similar observation can in fact also be made for the world ship fleet.

The above discussion on the evolution of the air transport market and the development of aircraft productivity suggests a rather continuous development pattern. As a final example in this Section we will show that the evolution of aircraft propulsion technology did not follow a similar pattern, but is instead characterized by an envelope, consisting of two successive growth pulses, with a period of saturation and discontinuity in between. Figure 3.4.7 reports an analysis of the evolution of the performance of aircraft engines [maximum rated horsepower (hp) for piston engines, and maximum take-off thrust in kilopond (kp) for jet engines].

Figure 3.4.7 shows the improvement in the performance of civil aircraft engines since the beginning of aviation. The first piston engine plotted on Figure 3.4.7 is the French 50 HP Antoinette engine introduced in 1909, and the last one is the American Wright Turbo Compound rated at 3,400 hp (i.e., at 90 percent of the estimated saturation level in piston engine performance). This last generation of piston engines, when approaching their ultimate performance limits, represented – even by today's standards – highly elaborate achievements in engineering, including double ("stretched") cylinder rows, four valve engines, and compound turbocharging to increase shaft power. These engineering developments after some 50 years, are now entering into the automotive industry. This could be an indication that the technological development of the internal combustion engine

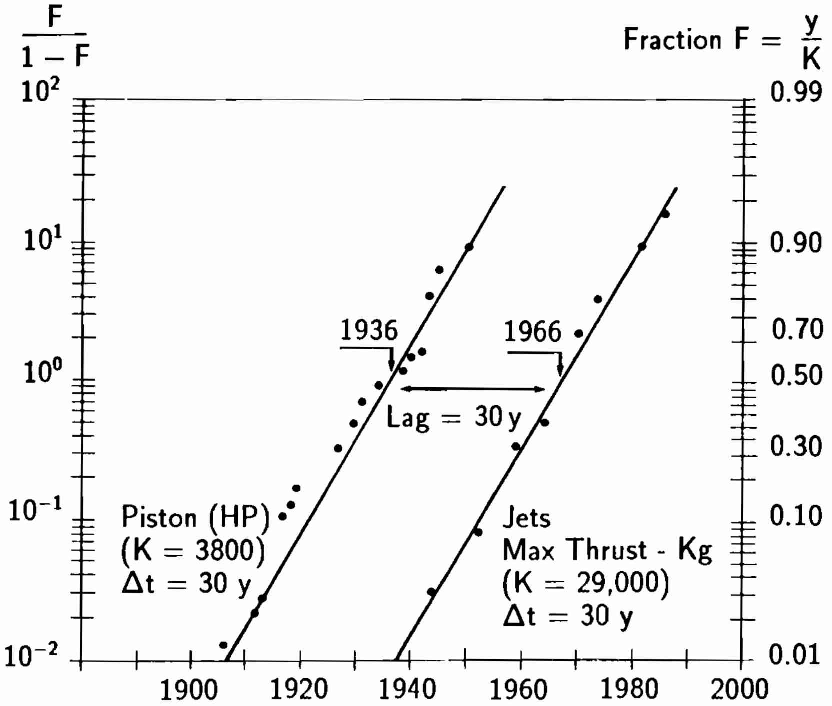


Figure 3.4.7. Performance levels of aircraft engines (pistons: K 3800 HP; t_0 1936; Δt 30 years; jets K 29,000 kp; t_0 1966; Δt 30 years). (Data source: Grey, 1969, Angelucci and Mantricardi, 1977, and Jane's, 1984.)

of cars is approaching a similar performance feasibility limit as aircraft piston engines were approaching in 1940. In fact, SAAB Scania has recently announced the development of a turbo compound diesel truck engine.

The development of piston engines is followed (with a lag of 30 years) by a parallel growth pulse of jet engines, starting with the German Junkers Juno 004 (rated at 900 kp) and ending with the American Pratt and Whitney JT9D of the early 1980s with around 90 percent of the estimated thrust saturation level. Both pulses are characterized by similar time constants (Δt) of 30 years, with the two inflection points being equally spaced at 30 year intervals. These midpoints in the development of aircraft propulsion technologies coincided with the introduction of the DC-3 aircraft and (approximately) the B-747 respectively. The parallel development of two successive technological generations of aircraft engines is a striking finding in itself which would merit further research. It appears that the development of the jet engine is best interpreted as follows: The technical feasibility limit reached in the development of the piston engine was overcome by the development of a new technology.

Thus, certain parallels in the dynamics of the evolution of passenger aircrafts and of aircraft power plants are evident, beyond the obvious similarity in their time constants (Δt of 30 years). The development of aircraft propulsion technology did not evolve continuously as did the performance of aircrafts and the market expansion of air traffic in general, but through two development pulses characterizing two successive generations of aircraft propulsion technology. Nevertheless these streams of development evolved consistently and synchronously. The approaching saturation in the growth of the performance levels of jet engines is an additional indication of the gradual improvement of aircraft technology development over the next 20 years. Future developments in jet engines will thus consist mostly of further improvements in fuel efficiency (like the Ultra High Bypass engine designs resulting in a 40 percent fuel efficiency improvement) and noise reduction and not in further increases in the rated power of turbines. As with the "stretching" of the size of the B-747, the development of piston engines, after having reached their inflection point in 1936, was also achieved by "stretching" (i.e., doubling of cylinder rows). A good example of this is the Pratt and Whitney Twin Wasp powering the DC-3 when introduced in 1935.

The time constants in the development of civil aviation and aircraft technology are thus around 30 years. Thirty years after the standard industry design emerged symbolized by the DC-3, the B-747, as the first wide-body jet, entered service. Its productivity represents about half of the estimated industry saturation level that might be approached around the year 2000. Thus, the life cycle of aircraft development spans about 60 years in total, from initial standardization and subsequent rapid growth, characterized by a large number of successive models introduced into the market, through to the inflection point of development when emphasis begins to shift to gradual improvements and increasing competition, characterized by cost reductions and rationalizations. Based on our analysis, the technological (performance) development of aircraft should approach saturation around the year 2000, with the present growth pulse in the world air traffic volume entering a phase of saturation some time thereafter. This, of course, raises the question about what may happen after such a saturation phase. Could one expect a new growth pulse (as we would be inclined to believe), a period of stagnation and subsequent decline (similar to the railways after the 1930s), or a period of instability with changing periods of market growth and decline?

After the saturation of the present growth pulse

The fastest airliner, Concorde, and the most productive jet aircraft, the Boeing B-747, flew for the first time almost twenty years ago. Ever since that time the volume of all airline operations has continued to increase while the productivity of the B-747 increased marginally only through stretching. Most of the subsequent technological development focused on gradual improvements such as noise reduction and the improvement of fuel efficiency. Unavoidably, this implies diminishing returns in the further advancement of airliners and, consequently, the possibility of an approaching saturation in commercial air transport.

Why should one then expect a subsequent further growth pulse in the volume of world air traffic after the year 2000? Marchetti, 1987, argues that the world economy and our societies are heading towards true globalization, which would effectively be achieved if any point on the globe could be reached within the daily time budget constraint of around one hour. Taking this a step further, one could imply that future developments in communication technologies could transform our globe into a "world village". Increasing global integration of economies and emerging new industrial production organizations with different structures of integration (horizontal as opposed to vertical as argued by Pry, 1988) and specialization in the production process (increasing importance of "software", highly specialized "custom design", and real-time production of industrial products, as discussed by Ayres, 1988) could be part of an emerging new paradigm for economic growth. Both tendencies would, as in the past, have important consequences in creating specific demands for passenger and goods traffic.

If such emerging tendencies would indeed become predominant features of the future functioning of our societies and economies, new transport systems with significantly higher performance levels, both in terms of transport speed and throughput, would become necessary. It is unlikely that significant increases in the number of aircraft in use or the development of hyperjumbos seating a few thousand passengers will occur, considering current airport overcongestion and the many inconveniences associated with an increased size of aircraft. Another possibility would be to enhance the productivity of the fleet by increasing aircraft speed.

Higher transport speeds, say in the range of Mach 5–6 and ultimately around Mach 25, would be required in order to integrate the whole globe into a functional entity like a present day urban agglomeration. Along similar lines, large *functional* urban agglomerations (i.e., in the form of a "megalopolis" or ultimately in the form of an "ecumenopolis", i.e., an urban cluster ultimately spanning a whole continent, Doxiadis and Papaioannou, 1974), transgressing our present day definition of cities in terms of their administrative boundaries, could only become possible by high speed and capacity transport infrastructures. The first indications of the development of such "functional" city corridors can be observed for instance in the USA (Boston–New York–Washington) or Japan (Tokyo–Osaka, with a total population of nearly 100 million people). It is not incidental that these large city corridors are connected by aircraft links operating like shuttle buses (in Japan by specially modified B-747s for the short distances travelled) or by the predecessors (Japan's Shinkansen or the French TGV) of new high speed and throughput ground transport systems. In future, regional links served by Maglevs, for instance in the form of long-distance metros, as for example proposed for Switzerland (*Die Presse*, September 9, 1988), could complement longer distance air connections. Under such an assumption, one could imagine two types of high performance transport infrastructures. High speed regional links integrating whole urban corridors into single functional cities, either by high speed train or metro type connections (like the planned new Shinkansen Maglev), or by specially designed medium-range aircrafts [as for instance developed at present by Messerschmitt-Bölkow-Blohm GmbH (MBB) for China]. The highest (spatial) hierarchy of exchange relationships could in turn be provided by a synergistic interlinkage of new communication and information exchange infrastructures, complemented by a new generation of high speed (supersonic or even hypersonic) aircrafts serving a limited number of transcontinental hubs (like the off-shore Kansai airport in Japan).

As the major productivity indicators (speed and capacity) of current aircraft technologies are apparently approaching upper limits, further growth in air transport demand would require new technologies that would extend the feasible and economic performance envelope of passenger aircrafts to span subsonic, supersonic, and eventually even hypersonic flight regimes. The most daring of the proposed designs could fulfill even more ambitious objectives and lead to the development of an air-breathing spaceplane. Currently such technologies are under active development in a number of countries, including the USA, France, Germany, Japan and the USSR. The basic objective is to develop new types of aircrafts that could extend the ultimate productivity limit of conventional aircraft (1.2 million passenger-km per hour). This will certainly be possible with aircraft speeds that significantly exceed the sound barrier.

Assuming that such a development is technologically possible and economically viable, it poses a number of fundamental questions associated with the expected side-effects of a large fleet of hypersonic passenger aircraft. Certainly, the noise problem must be overcome both by reducing the inherent noise profiles and removing the hubs and routes for such aircraft from populated areas. A more important concern, however, is the possible adverse effect of such a fleet on the chemistry of the upper atmosphere. In view of the observed ozone depletion of the stratosphere, such concerns are timely, while the design and configuration of proposed hypersonic transports are still in an early development phase.

The most likely designs for such new types of aircraft would favor alternative fuels, since conventional jet fuel becomes impractical at speeds exceeding around Mach 3.5. Here methane (natural gas), and at speeds above Mach 5-6 hydrogen, would be the fuels of choice, since, apart from other reasons, these fuels would create the most environmentally benign emissions (Victor, 1989). New advanced types of aircraft could in fact be indicators of the forthcoming increased importance of these energy carriers. Increased use of natural gas (methane) could prepare the ground for a hydrogen economy in the more distant future in terms of the related infrastructure for transport, storage, and use of gaseous energy vectors.

At present it is of course difficult to assess a scenario that would enable a truly world economy. If the possible adverse effects of a continued increase in transport operations could, however, be overcome, the further maximization of range through productivity increases and speed by a new generation of transsonic aircraft, would be consistent with the observed evolution of transport systems since the onset of the industrial revolution. For the first time in our history, it might become possible to mesh the world's principal gravity centers into a single functional entity. The availability of a new air transport technology could therefore lead to a new growth pulse in air traffic, increasing traffic flux levels at least by one order of magnitude* above our estimated saturation level of world air transport based on present technology.

Finally, one should also recall, that at the time of the large transatlantic liners their travel speed (the fastest connection by the Queen Elizabeth II took around 5 days and 4 nights between New York and Southampton) was considered largely adequate by contemporary sources. Based on this opinion, the market potential for transcontinental air passages should have been assessed as being rather limited. However, as shown above (Figure 3.4.1), the market response was rather different. Since the first non-stop flight from Europe to the USA and back (with a Lufthansa Focke-Wulf Fw 200 "Condor"

---

* This hypothesis is based on an analogy of the market impact of the introduction of the jet aircraft.

plane), air traffic has been developing rapidly over the Atlantic, and in 1987, the last regular transatlantic passenger liner, the Polish Stefan Batory, discontinued service.

Summary

Figure 3.4.8 summarizes the various developments in aircraft performance, air traffic volume, and aircraft propulsion technology discussed so far. It becomes clear, that the growth associated with the aircraft technology cluster will continue for some time after the saturation period of the automobile/oil technology cluster discussed in Section 3.3 above. Thus, air transport and aircraft technology provide an example of a technological cluster "tunneling through" the saturation period of the dominant technological (automobile/oil) regime. Such was the case with railways during the saturation phase of canal development, with cars during the saturation of the railway expansion pulse, and today it could be the case with air transport in the approaching saturation phase of car motorization.

We have discussed above the relationship of new aircraft technologies to new energy vectors required for fueling high speed aircrafts: methane up to speeds of Mach 5-6, and hydrogen for even higher speeds. We complement our summary by providing an indication of the energy technology which could also "tunnel through" in its growth phase, the present saturation of the market dominance of oil. Natural gas appears to be the energy carrier, which could well provide the potential for gaining further market shares in the future, eventually to become the single most important source of primary energy after the year 2000 (see Hefner, 1984, Rogner 1988, and Grübler and Nakicenovic, 1987). This is illustrated by examples from energy transport infrastructures. The transport infrastructure of coal and its saturation in the 1930s was already discussed in the Section on railways. Along with the present saturation of the share of oil in the primary energy balance, we observe an apparent saturation of oil related transport infrastructure (pipelines) in the USA. Contrary to this, there still appears to be a growing tendency in the importance (measured by the size, i.e., the length of the pipeline network) of natural gas transport infrastructure in the USA and indeed at the global level too. Both aircraft and methane technologies could thus "tunnel through" the *season of saturations* of the oil/internal combustion technological cluster, responsible for much of the economic growth of the post World War II period.

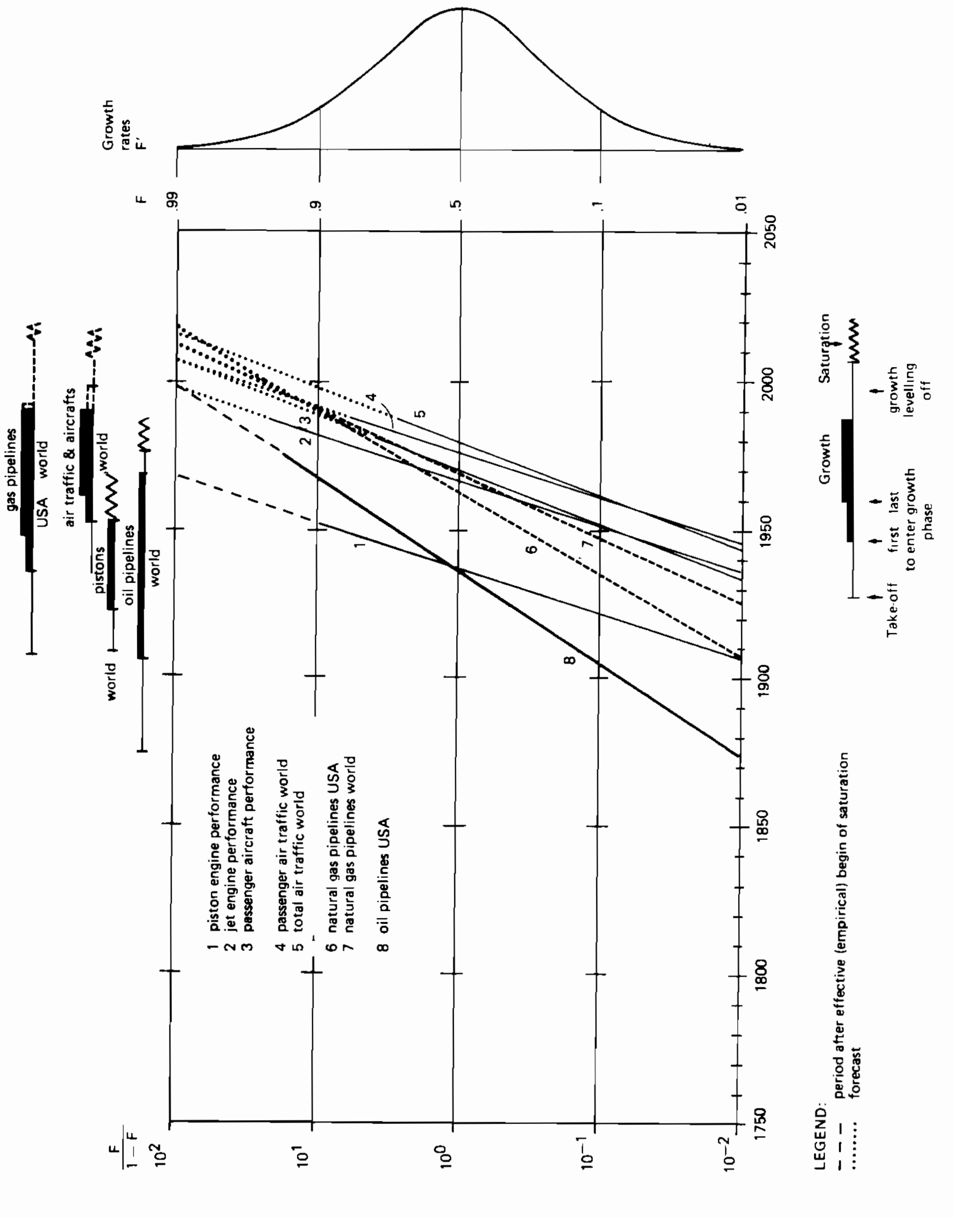


Figure 3.4.8. Evolution of air traffic, aircraft technologies and of energy transport infrastructures (oil and natural gas). (Source: Passenger aircraft performance and passenger-km transported worldwide: Nakicenovic, 1988, oil and natural gas pipelines in the USA: Grübler and Nakicenovic, 1987, natural gas pipelines at the global level: Marchetti, 1987, all others: author's own estimates.

CHAPTER 4

An Integrative View of Long-Term Infrastructure Development

Evolution is the result of a sequence of replacements.

Elliot W. Montroll, *Social Dynamics and the Quantifying of Social Forces*, 1978

In the previous Chapter we have sketched out a quantitative history of individual transport infrastructures and of technological change in the devices using them. In order to explain some of the historical development patterns, reference to other transport infrastructures and technologies was made in discussing the whole life cycle of an infrastructure system, i.e., not only its growth, but also its subsequent saturation and eventual decline. The purpose of this Chapter is thus twofold. First we attempt to integrate the various evolutionary tendencies in the development of individual infrastructures into a holistic view of the whole transport sector. For this purpose different measures are used to assess the relative importance of individual transport systems at various times in history. First, the changing *structure* of the whole transport infrastructure is analyzed using both physical length and the value of the capital stock as an indicator (Section 4.1). Secondly the relative *output* of different infrastructures in terms of the passenger-kilometers (km) and ton-km performed (i.e., the modal split) is analyzed (Section 4.2). We attempt to show, that from a long-term perspective, the evolution of the transport system both in terms of its structure as well as its performance (modal split) is characterized by a regular sequence of replacements of older forms by newer forms in order to satisfy human transport needs. These long-term trends appear to be robust even against major societal and economic disruptions and discontinuities.

The second objective of this chapter is to develop some generalized propositions as regards the driving forces behind the evolutionary transformations in the transport sector (Section 4.3). These driving forces are seen in the *performance* (in particular *speed and range*) and in the *service* (flexibility and quality) characteristics of different transport systems, and how they respond to the continually evolving societal requirements. In a subsequent step, we discuss some economic aspects of the long-term evolution of the transport sector. We argue, that the (price insensitive) high-value market niches in the transport sector, i.e., moving people, information, and high-value goods, are of vital importance in the long-term development of an infrastructure. It is these market niches that are conquered first by a new competitor and consequently the ones the older systems tend to lose first. Only at a later stage, when technological and engineering improvements, and market growth result in a significant reduction in transport costs do new infrastructure systems also enter into lower value market segments, finally to become themselves an indispensable (price regulated) "commodity".

We then consider the long-term relationship between economic activity and transport requirements (Section 4.4). This analysis suggests, that we may be in a period of structural transition in the material intensity of highly developed economies. For the first time since the onset of the industrial revolution, the transport requirements per unit of value generated, are not growing but seem to be on the decline. This is the result of a structural shift in the output of developed economies in the direction of less material intensive (lighter) and considerably higher value goods. If this tendency should continue in the future, important repercussions on the quality of the transport system (especially as regards speed, flexibility, and quality of service) will result. This is consistent with our analysis of the relationship (elasticity) of the modal split in goods transport to the value of goods transported. On the other hand, no such observation can be made with respect to passenger traffic, which is expected to continue to grow at significant rates. Certainly in the area of air transport, this is possibly not yet fully realized (in terms of potential bottlenecks in the related infrastructures) by transport policy makers and planners.

This leads to the final hypothesis with respect to the driving forces behind the long-term evolution of the transport system: the relationship between transport and communication (Section 4.5). Seen from a long-term (200 years) perspective, transport and communication appear to evolve in unison, and to be of a synergistic, cross-enhancing nature. Transport and communication are seen consequently as *complementary* rather than *substitutable* economic goods. From such a perspective, it appears again unlikely that we are in a phase of structural discontinuity in the growing trend of passenger transport. New infrastructures and organizational and institutional settings will be required to provide fast, safe, reliable, and functionally (as well as aesthetically and environmentally) adequate transport services for future increases in travel demand.

4.1. The Structural Evolution of Transport Infrastructures

In the previous analyses, the USA has always been used as a prominent example when discussing the quantitative evolution of its transport infrastructure. The reason for this was due to the unique data availability with respect to long time series on infrastructure development (including for instance data on surfaced roads which are almost entirely lacking in other countries). Consequently we start our integrative discussion on the structural evolution of the whole transport infrastructure system with the case of the USA. Later, we discuss the case of Canada and the USSR, which are the only other countries where data availability permits a study of the long-term evolution of *all* transport infrastructures (canals, railroads, surfaced roads, and airways).

USA

Figure 4.1.1 summarizes the history of individual transport infrastructures, measured by their respective length, in the USA. An analogy, with data* on air transport infrastructures (the functional analogy being already contained in the semantics of the term *airways*), complements the quantification of *physical* infrastructures.

The length of all four transport infrastructures has increased five orders of magnitude since 1800. Each successive transport infrastructure expanded into a network ten times larger than the previous one. In addition, it is interesting to note that, measured by length, new

---

* Source: US DOC, 1975, and 1976 to 1985. As a measure of the extension of the air transport infrastructure, we use the total route mileage operated by airways. As this figure was no longer available for years after 1976, these were estimated based on the growth of Federal airways during the period 1977 to 1980. For years prior to 1930 the estimates from Woytinsky, 1927, have been used. Our estimate is of course a crude measure, as it does not describe any qualitative characteristic (regularity of service, frequency, and throughput of air transport) of the air transport being carried out on these routes. For our purposes here, it serves as a rough indicator of the growing importance of air transport infrastructure, although it may not be directly comparable with the physical, dedicated infrastructures for water, rail, and road transport.

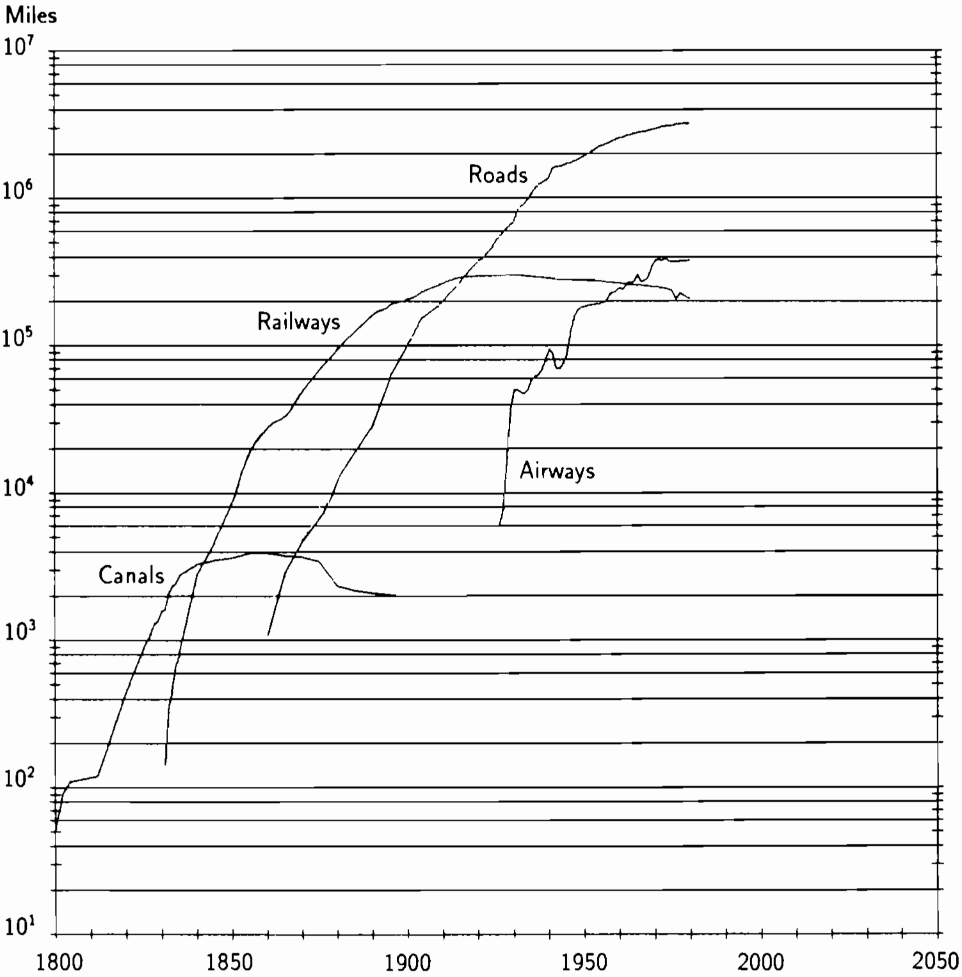


Figure 4.1.1. Evolution of the length of transport infrastructures (canals, railways, surfaced roads, and federal airways) in the USA (in 1,000 miles).

infrastructures overtook existing ones only at the time the latter started saturating. This was the case with canals and railways in the 1840s, and also with railways and surfaced roads prior to the 1920s. Based on this historical pattern, one would expect airways to become a dominant transport infrastructure only after the expansion of the road network is completed.

Figure 4.1.2 summarizes the expansion of the three *physical* infrastructures in the USA discussed in detail in Chapter 3, normalized with respect to their respective saturation levels (by plotting the relative length as a percentage of the saturation level). The succession of individual infrastructures can be described in terms of three S-shaped

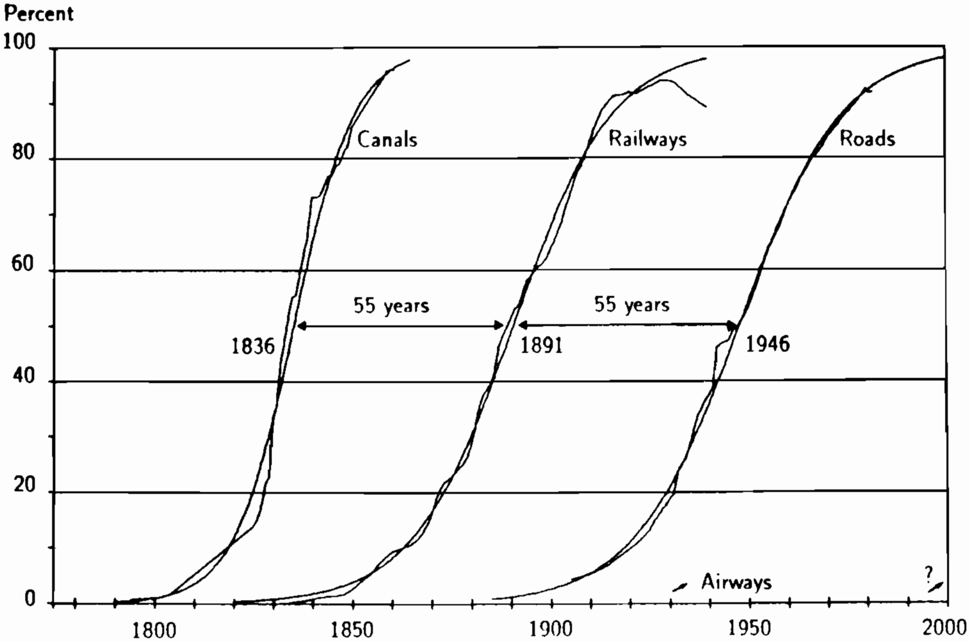


Figure 4.1.2. Growth to limits of canals, railroads, and surfaced roads in the USA. Note that absolute saturation levels are different between the three infrastructure systems.

growth pulses that are given together with the estimated logistic curves.

The development of canals, relative to the saturation level achieved, was much quicker than the expansion of railways and surfaced roads. The time it takes to grow from 10 to 90 percent of maximum network size, (Δt) is about 30 years for canals, 55 years for railroads, and 65 years for surfaced roads. The midpoints between the individual infrastructure growth pulses (i.e., the time of their maximum growth rate) are spaced 55 years, as are their periods of saturation. It is remarkable that the saturation and onset of decline of all three infrastructures coincides with prolonged economic recessions (i.e., in the 1870s, 1930s, and 1980s). At the same time these periods of structural discontinuity see the emergence of new transport systems: surfaced roads around 1870 and air transport in the 1930s.

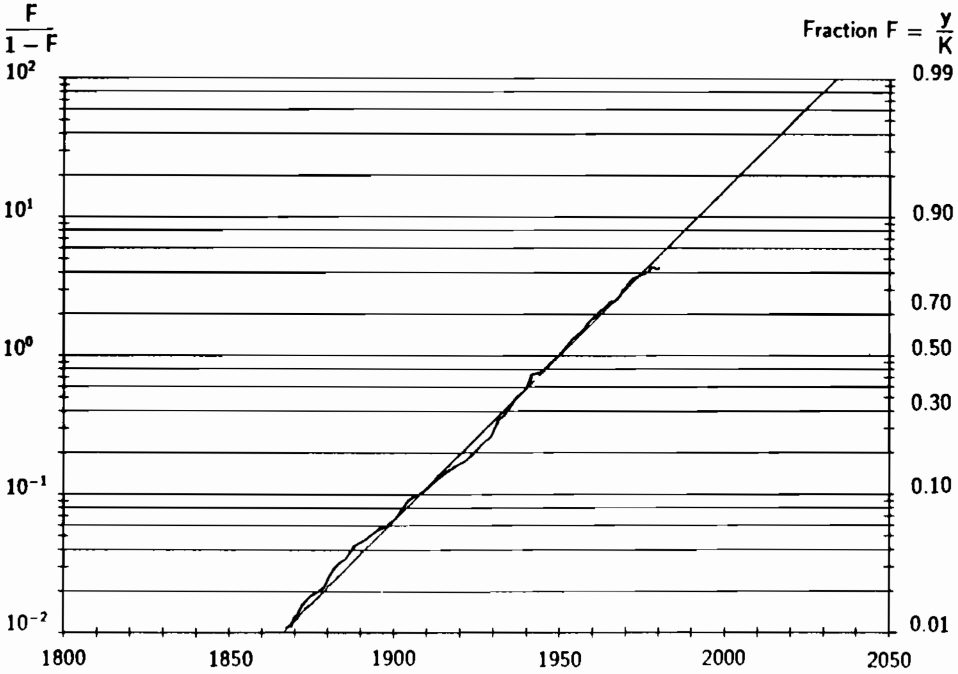
The "growth to limits" to different absolute levels and the subsequent onset of decline of individual transport systems gradually changed the structure of the transport infrastructure system and in particular, the relative importance of individual networks. This sequence of replacements in the relative importance of individual infrastructures already points to a possible analogy to market substitution. Seen from such a perspective, we may interpret the successive sequence of "life cycles" of individual infrastructures as a change in the morphology or structure of the transport infrastructure network of the USA. Previously dominant infrastructures lose out in importance to newer systems, which in turn tend to saturate themselves and become replaced by newer infrastructures.

Figure 4.1.3 illustrates the share of individual transport infrastructures in the total length of the transport system of the USA. In addition, we depict the evolution of the length of the total transport infrastructure of the USA, i.e., the expanding "niche" in which individual transport infrastructures "compete" for relative positions with respect to the length of their networks. An interesting finding of our analysis is that when the different transport infrastructures, including airways, are regrouped together, they appear to grow according to a logistic growth pulse depicted in the upper part of Figure 4.1.3, with a Δt of around 80 years and an estimated ultimate saturation level of some 4.7 million* miles (around 7.6 million km). Thus, despite the fact that individual transport infrastructures have evolved through a series of discontinuities, marked by a sequence of introduction, logistic growth, subsequent saturation, and finally decline, the *aggregate* in measuring the total length of all transport infrastructures (i.e., the envelope resulting from the growth and decline processes of the individual infrastructures) appears as a smooth regular growth process. By 1980, it appears that 80 percent of the ultimate saturation level in the growth process was reached. Also, it was only by the 1860s (the saturation date for the canal network expansion) that the one percent mark of the estimated final saturation level was exceeded.

The apparent "homeostasis" in the growth of the transport infrastructures of the USA is noteworthy in that the saturation and later on, decline of particular individual networks (as in the case of canals and railways) was up to now "filled" by the growth of newer infrastructures, consistent with the logistic "envelope" of Figure 4.1.3. This is also frequently observed in the evolution of dynamic, self-organizing systems in chemistry or biology. The growth of this "envelope", proceeds with a Δt of 80 years slower than the growth of individual infrastructure systems (Δt of canal growth around 30 years, railways 55 years and roads around 65 years) and saturation, if this process should continue its historical pattern, would occur around the year 2030.

---

* We estimate a 90 percent probability that the final saturation level K will be between 4.4 to 5.0 million miles. The R^2 of estimate is 0.9978.



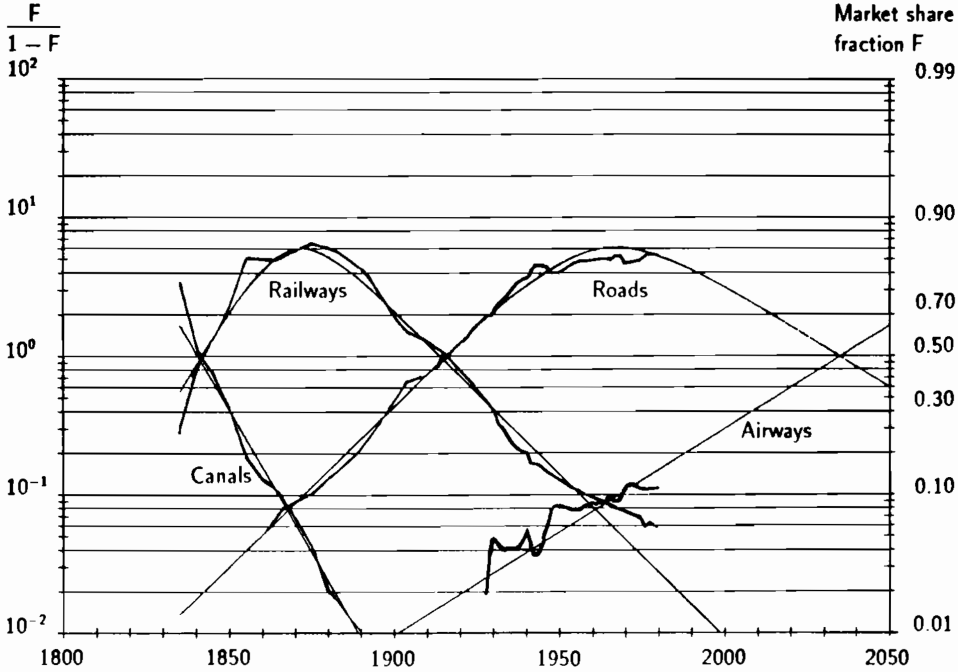


Figure 4.1.3. Growth in length of total transport infrastructure [K 4.7 million miles; t_0 1950; Δt 80 years] (top), and infrastructure substitution (bottom) in the USA.

The bottom of Figure 4.1.3 presents the structural changes in the transport infrastructure of the USA, organized with the help of a multi-variate logistic substitution model. Despite the fact that the model fails to describe the rapid early phase in the introduction of the railways (i.e., prior to 1835), and the fact that the growth of air transport infrastructure proceeds rather turbulently, the regularity of the sequence of replacements in the importance of individual infrastructures is striking.*

This particular model representation shows the relative importance of competing infrastructures and the dynamics of the structural evolution process along a sequence of replacements over the last 160 years. In any given period, there is a clear market dominance (i.e., more than a 50 percent share) and at the same time a simultaneous spread of transport activities over two or three different systems. Thus, while competing infrastructures are all simultaneously used, their mix changes over time. Three main periods in the structural changes of the transport infrastructure of the USA can be distinguished.

- Growth of the railway network (main tracks operated) and resulting decline in importance of canals as a transport infrastructure, with a Δt of around 45 years. Note here, that around 1860 (maximum size of the canal network) canals accounted for only more than 10 percent of the total transport infrastructure (canals, railways, and surfaced roads) of the USA.
- Saturation of the importance of railways (in their share in the length of transport infrastructures) around 1870 (i.e., around 50 years prior to the time when railways reached their maximum network size) and subsequent decline in importance due to the growth of surfaced roads proceeding with a Δt of around 80 years.
- Long saturation phase (from around 1940 to 1980) during which roads constituted the most important transport infrastructure in the USA, however, losing out in the long-run to air transport infrastructure, growing with a Δt of around 130 years.

---

* The turbulent trajectory of air transport infrastructure should not be surprising, as we deal here with a conceptual and functional analogon, compared to *physical* infrastructures. Thus, whereas canals, railroads or surfaced roads evolve through a stable trajectory as a result of the inherent long lead times and costs in building up and decommissioning physical systems, air routes may easily be opened and closed down following particular market constraints (see e.g., the turbulent period during World War II in Figure 4.1.3). Despite the fact that we use a different data set, the estimates with respect to airways in the infrastructure substitution in the USA are very close to the tentative results reported in Nakicenovic, 1988, and support the resulting conclusions.

Several invariant features in the long-term structural evolution of the US transport system emerge from Figure 4.1.3. The dominant (longest) transport infrastructure is always longer than at least half of the total length. Consequently, the second and third longest infrastructures account for less than half of total length. This symmetry in the dominating role of individual infrastructures (canals prior to 1840, railways between 1840 to 1920, and roads since the 1920s and possibly well into the 21st century), is complemented by another phenomenon: the time constants Δt in the growth (decline) of importance in the total transport infrastructure length increases from around 45 years (decline of canals and growth of railways) to 80 years (decline of railways and growth of surfaced roads), and appears to increase further to some 130 years (decline of roads and growth of air transport infrastructure). This could be the result that each new infrastructure system expands to a larger network size than the previous system (although the Δts of the individual expansion pulses increase only from 30 to 55, and 65 years for canals, railways, and roads respectively). A corollary to this is that it apparently takes longer and longer to phase out transport infrastructures. Could it be that in the very long-term we are heading towards a situation of maintaining all infrastructure systems (as a kind of insurance policy), or are we approaching increasing levels of specialization in the services rendered by different transport infrastructures? Finally, we note that the distance between the maxima in the share of total infrastructure length between railways and surfaced roads is about 100 years, indicating the considerable time span involved in the transition from the dominance of one infrastructure system to the next one.

The difference in the dynamics (Δts) of the growth of individual infrastructures and of their relative shares in total infrastructure length, might appear at first sight to be a contradiction. However, this difference stems from the complex coupled dynamics of total infrastructure growth, and the growth and decline rates of individual transport infrastructures. As the total length of infrastructures increases, even the rapid growth of individual infrastructures such as airways, will translate into slower growth rates in their *relative shares*. Once, the growth rate of an individual transport infrastructure falls behind the growth of the total system, its relative share starts to decline. Thus, the share of railways in total infrastructure length decreased as of 1870, whereas railways continued to expand until the end of the 1920s. Similarly, the length of the surfaced road network still continues to increase, albeit slowly, as it is close to apparent saturation. However, its relative share already started to decrease in the 1960s. Thus, the total length of an individual infrastructure (in this case canals, railways, and surfaced roads) can still be growing, even decades away from ultimate saturation and subsequent senescence in absolute network size, while its share in the length of the whole transport system is already declining. Therefore, the saturation and decline in the relative market share precedes the saturation in *absolute* growth in an expanding market. This implies that the eventual saturation of any competing technology may be anticipated in the substitution dynamics of a growing market, such as for railways as early as 1870 and for roads as of 1960. The infrastructure substitution model measuring market shares, may thus be considered to be a "precursor" model, in the long-term life cycle of individual infrastructures.

Let us conclude by stressing the regularity in the rise and fall of the importance of individual transport infrastructures. This regularity neither appears to be affected by the discontinuities in the development of individual transport infrastructures as part of their proper life cycle (growth, saturation, and subsequent decline), nor by external events like the depression of the 1930s or the effects of major wars. We surmise that this stable behavior is the result of an invariant pattern in the societal preferences with respect to individual transport infrastructures, which result from differences in the performance levels (seen as a complex vector rather than represented by a single measure) inherent in different transport infrastructures and technologies.

At this point the reader might object to our conclusions by pointing out that the stable evolutionary pattern described above may be a specific isolated case for the USA. A second, justified objection would be that our analysis of the long-term infrastructure substitution was based on a rather crude measure of the length of different infrastructures, ignoring therefore the different costs for construction and the different capacities (throughputs) and utilization rates realized in different parts of the transport infrastructure network.

Below we show that the pattern observed in the case of the USA is invariant, in that a similar structural evolution can also be observed in the case of a planned economy (the USSR). In the case of Canada, instead of length, we use a monetary indicator (quasi capital stock) to assess the relative importance of individual infrastructures. We also analyze below the *utilization* (*throughput*) of different infrastructures, i.e., the modal split (passenger- or ton-km transported) of the whole transport system and conclude, that the dynamic picture outlined above, is confirmed by using alternative indicators.

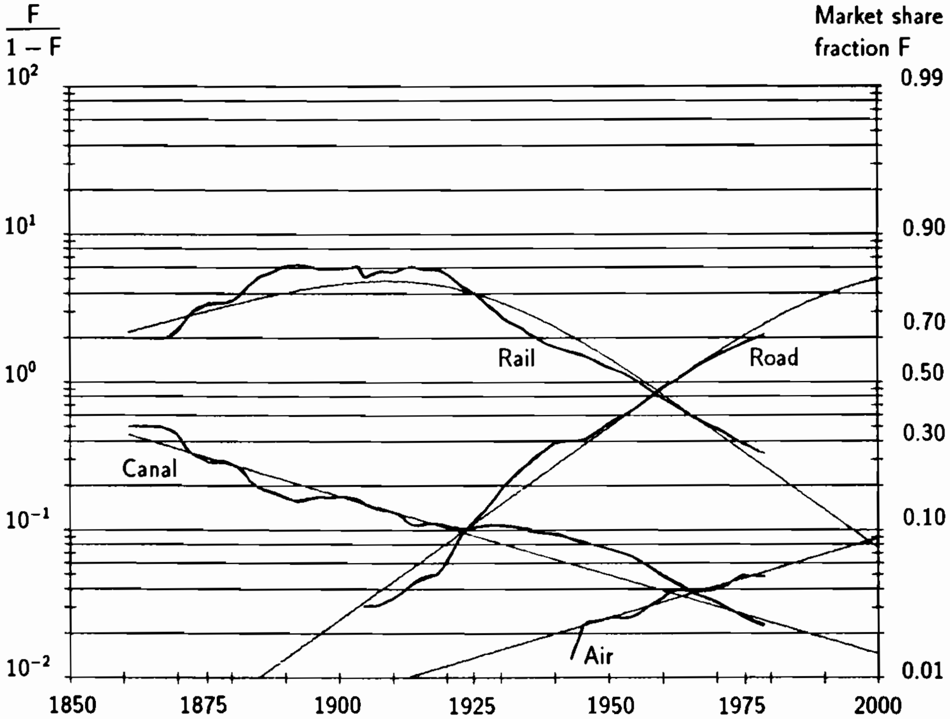


Figure 4.1.4. Infrastructure substitution in Canada, in fractional shares of canal, railway, road, and air transport infrastructures in quasi capital stock of the transport system, measured in constant 1980 Canadian dollars.

Canada

Figure 4.1.4 deals with an analysis of the changing structure of the transport infrastructure in Canada. The share of different infrastructures in the capital stock* is used as an alternative measure of their relative importance. The basic pattern in the sequence of replacements identified for the USA, is also confirmed by this analysis. However, the relative importance of air infrastructure in the capital stock is, due to its "immaterial" nature, smaller than its relative length or its share in the output of the transport sector.

It is even more important to realize the very long time constants involved in the structural transition in the composition of the transport sector capital stock: the Δts for the decline of railways and

---

* The quasi capital stock (excluding depreciation) is measured in constant Canadian 1980 dollars, derived from cumulative investment data given in Haritos, 1987, and Historical Statistics of Canada, 1965. Railway investment data prior to 1850 were based on marginal railway construction cost estimates using data from Urquhart, 1986.

growth of road infrastructure are in the order of 80 years, and the decline of canals and growth of air infrastructure even slower with Δts of nearly 200 years. Transport structures are built to last, they are almost immortal. Once their original use declines in importance because of changing societal preferences and requirements for the transport system, new uses may evolve. Two good examples are the leisure activities that have evolved around former canals, e.g., the Rideau Canal in Canada, and the 19th century museum opened in Paris in the former Gare d'Orsay, an impressive monument of the railway era.

USSR

Figure 4.1.5 shows the substitution of transport infrastructures in the USSR and Tzarist Russia before the revolution which is analogous to the analysis of the USA shown in Figure 4.1.3. As for the USA, the analysis includes also airway* infrastructures in order to calculate the relative shares of different systems in the length of the whole transport infrastructure of the USSR. Despite the fact that there may be some uncertainty related to our estimate of the extent of air transport infrastructure in the early days (pre-1965), we observe in both countries a similar dynamic picture with respect to the structural changes in the transport infrastructure. Again the pattern of temporal changes is marked by a high degree of regularity and the quest for higher speed and productivity.

The similarity in the pattern of structural change in the two countries is noteworthy for several reasons.

- It demonstrates a continuity in the structural shifts of the transport system which does not appear to be affected by major societal discontinuities such as the October Revolution.

---

* As a measure of airways length, the total length of airways of union importance is taken (Lewytzkyj, 1979, and Narod. Khoz, 1982, and 1987). We have used the data available as of 1965. Earlier years are estimated based on interpolation from data given for the 1920s in Woytinsky, 1927. The only difference between our estimate and official statistics is for the year 1940, where we estimate some 14,000 km of airways of importance for civilian air traffic. This is in contrast to the official statistics of over 51,000 km for 1940 (Narod. Khoz., 1982). This figure does not, however, appear realistic for two reasons: first in view of the low volume of air transport operations reported for that particular year (around 800 t-km passengers, mail, and goods transported per km airway, compared to 30,000 t-km per km airways in 1965 and currently over 55,000 t-km/km). Secondly, as our analysis concentrates on the civilian transport infrastructure, such a high figure for airways appears rather unlikely for 1940, i.e., a year when World War II spread to the USSR. Finally, one might refer here to the more general problem of data validity for statistics elaborated during the Stalinistic time period. In terms of our analysis the official figure for the share of airways in 1940 would amount to some 17 percent of total transport infrastructure length (as opposed to our estimate of 5 percent). Recall here, that airways reached the 20 percent level in the share of total infrastructure length in the official statistics only by 1965.

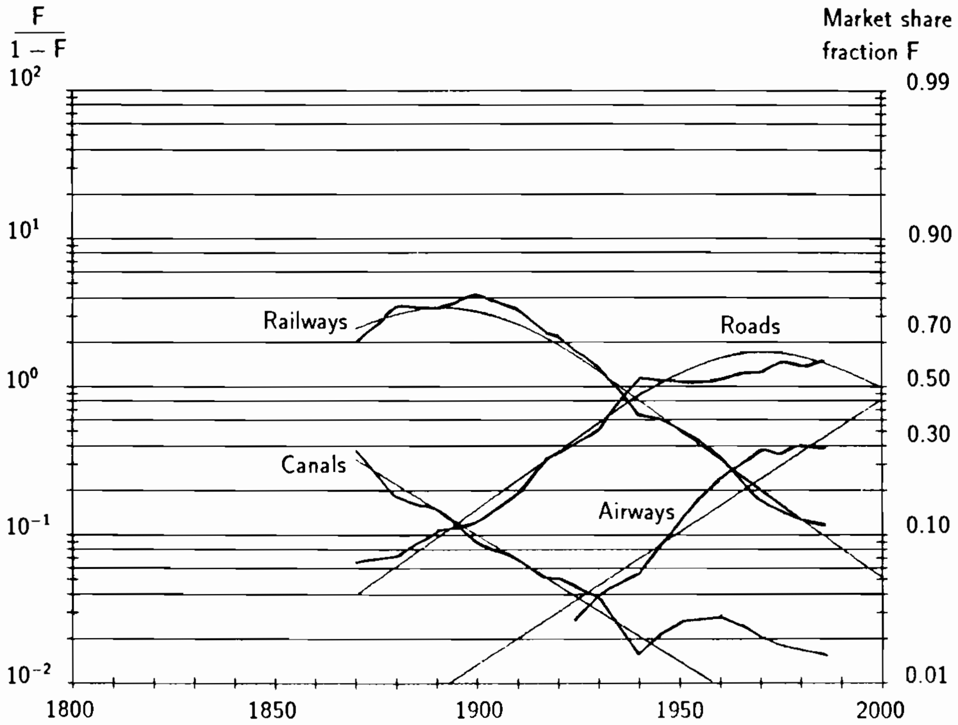


Figure 4.1.5. Infrastructure substitution in the USSR, in fractional shares of length of all transport infrastructures.

- The long-term dynamics and structural trends in the transport infrastructure development show similar results for both a planned economy and for a market economy.
- The long-term structural changes in the importance of different infrastructures in the USSR do not appear, in principle, to have been affected by the entirely different infrastructure development policy pursued by the USSR.

Recall here, that contrary to the USA, where the length of railways and canals is declining or stagnating at a low level, the USSR followed an ambitious canal and railway construction program. Furthermore, private car ownership in the USSR is at such low per capita levels (compared to the USA) that the dominance of road infrastructure in the transport system of the USSR may come as a surprise. We show below, however, that in terms of transport infrastructures, surfaced roads are of similar *functional* importance in the USSR, although road transport is not assured by private vehicles (cars) but by collective vehicles (buses).

Canals appear to have reached the end of their technological life cycle in the USSR. Despite their revival in the post World War II period, their share in the length of transport infrastructures has been hovering around the two percent market share level since 1940. Railways reached their maximum level of importance at the beginning of this century and since then have been losing in importance to the surfaced road network. The latter is presently saturating in terms of its relative share in the total length of transport infrastructures, to be replaced in turn by airways. The time constants of the long-term structural change processes are all very similar. The Δts for the replacement of canals and railways and of the growth (and long-term replacement) of roads and airways are all around 100 years, as is the spacing between the (estimated) peaks in importance of railways and roads in the total length of the transport infrastructure of the USSR (as in the USA). Thus it takes 100 years for the new transport infrastructure (roads) to take over the dominant role from the old transport infrastructure (railways). The similarity in time constants observed in the structural evolution of the transport system of the USSR, is different to that of the USA, where we have observed that each successive infrastructure system took a longer time constant to grow in importance in the total infrastructure length.

Over the last few decades, the structure of the transport infrastructures in the two countries has been converging. Figure 4.1.6 summarizes the infrastructure substitution pattern in the USA and the USSR. The structural differences between the two countries get smaller with each new transport system. The greatest difference is with canals, which phase out some 70 years later in the USSR than in the USA, and at a considerably slower pace. The decline of importance of railways proceeds at a similar rate in both countries, with the USSR lagging by about 30 years. This lag is also reflected in the phase shift of the peak of the importance of railways between the two countries (USA in 1870 and in the USSR about 30 years later). Surfaced roads approach saturation in both countries practically at the same time, although at a lower level in the USSR. Airways even appear to grow faster in importance in the USSR than in the USA. Currently, airways account for about one fourth of the transport infrastructure network length in the USSR compared to above 10 percent in the case of the USA.

However, for the year 2000 and thereafter we may summarize the infrastructure systems of both countries as follows. Canals will be almost extinct (in terms of relative importance), railways will be approaching the end of their technological life cycle accounting for only a few percent of the infrastructure length, whereas the dominant transport infrastructures will be roads with decreasing and airways

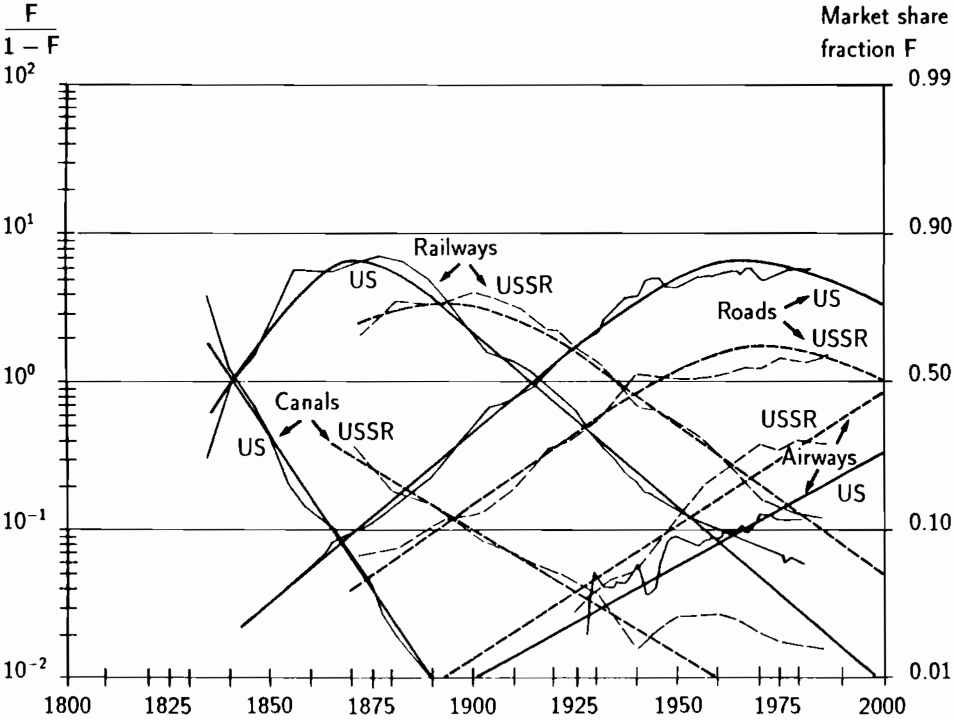


Figure 4.1.6. Infrastructure substitution in the USA and the USSR, measuring shares in length of all transport infrastructures.

with increasing importance. The similarities (although with some lagged developments as in the case of canals and railways) between the two countries point to a deeper causality for such similar structural changes in transport infrastructures. These causal forces appear disconnected both from the market clearing mechanism (market or centrally planned economy) and from the different relative transport cost/price structure prevailing in the two countries.

We assume that differences in inherent performance levels of individual infrastructures (in particular transport speed) and basic human decision criteria and preferences underlying the choice of different transport modes (such as individual travel time budgets), which transgress the differences in the economic systems of the two countries, are responsible for the similarity in development patterns. Finally, a further similarity in the dynamics between the two countries appears to exist in the area of the growth in the size of the total infrastructure network. Total transport infrastructure length in the USSR grows along a logistic trajectory with a Δt of 80 years (as in the USA), but lags by some 30 years behind the USA growth pulse (t_0 around 1980, compared to 1950 in case of the USA).

Thus, we conclude, that the historical evolution of transport infrastructures (measured by their share in the length or the value of the capital stock of the total transport system) is characterized by a regular sequence of replacements. New systems, that correspond better in performance levels to societal preferences with respect to transport infrastructures, replace existing ones. This evolutionary process appears invariant between different economic systems, and in addition is of a homeostatic nature. We believe that the main driving force responsible for this invariant evolutionary pattern relates to the human time allocation mechanism (the "law of constant travel time" as formulated by Zahavi, 1979, and 1981) favoring faster transport infrastructures. From this perspective, the growing importance of air transport and its infrastructure becomes more plausible, especially as we show in the next Section that air transport is of growing importance for moving the highest value "goods" over larger distances: human beings.

4.2. Evolution of the Performance of Infrastructures (modal split)

As a subsequent step in the analysis of the importance of various transport infrastructures, we analyze their relative contribution to the transportation output in different countries, i.e., the modal split in terms of passenger- and ton-km performed by various transport modes. As the analysis of infrastructure development concentrated on the evolution of long-distance transport systems, we consequently analyze the modal split for long-distance (intercity) passenger travel only. The analysis covers some shorter time series (since 1950) for the USA and the FRG as well as the case of the USSR, where available data permit us to go back as far as 1920. As another example we consider the case of France, where we have been able to reconstruct time series of total passenger modal split since the beginning of the 19th century. The emerging picture is rather turbulent due to the fact that the available data do not allow the separation of the two different markets for passenger transport, i.e., local and long-distance travel; France also experienced two major disruptions because of the two World Wars. Despite turbulences and the crude (aggregated) measure available, the results do confirm the analysis of the long-term tendencies of infrastructure development and passenger modal split, identified in the other countries analyzed.

Finally, we consider the case of goods transport, by discussing the modal split in France and Germany in terms of traditional tonnage-oriented measures. In addition we look at the modal split for goods transport in terms of *value* (imported or exported) for Germany and Sweden. The latter indicator appears more appropriate for describing the long-term changes in goods transportation, especially if the apparent transition to increased dematerialization and high-value goods production becomes a dominant feature in the future of advanced economies.

4.2.1. Passenger modal split

As a measure of the output (performance) of individual transport infrastructures we analyze for the following cases the long-distance passenger modal split. Being the highest value transport market niche, where competition between different transport modes is largest, long-distance passenger transport is a good indicator of likely future developments in the lower value market segments. The analysis reported below confirms the dynamic picture which emerged from Section 4.1: The long-term evolution of the transport system is characterized by a regular sequence of replacements, which appears invariant between different countries or even between different economic systems, pointing at much deeper underlying long-term driving forces than normally enter short-term transport demand and mobility models (such as relative transport price structure, private car ownership rates, etc.).

USA

Figure 4.2.1 shows domestic intercity passenger traffic in billion passenger-miles for the USA. The decline in the transport performance of railways, the stagnation of bus transport, the continued growth in car travel, and finally the rapid growth of passenger air transport, characterize the situation of this premium market segment of the transport sector. The model forecasts presented in Figure 4.2.1 are derived from the relative market share estimates of different transport modes (Figure 4.2.2) applied to (exponential trend) estimates of total market volume. Total intercity passenger traffic increased in the USA from 506 billion passenger-miles in 1950 to close to 1,820 billion passenger-miles in 1986, i.e., at an average annual growth rate of 3.6 percent. The different (positive or negative) growth trends in the relative performance (output) of different intercity transport modes in the USA since 1950 are quite accurately described by a logistic substitution model as presented in Figure 4.2.2 below.

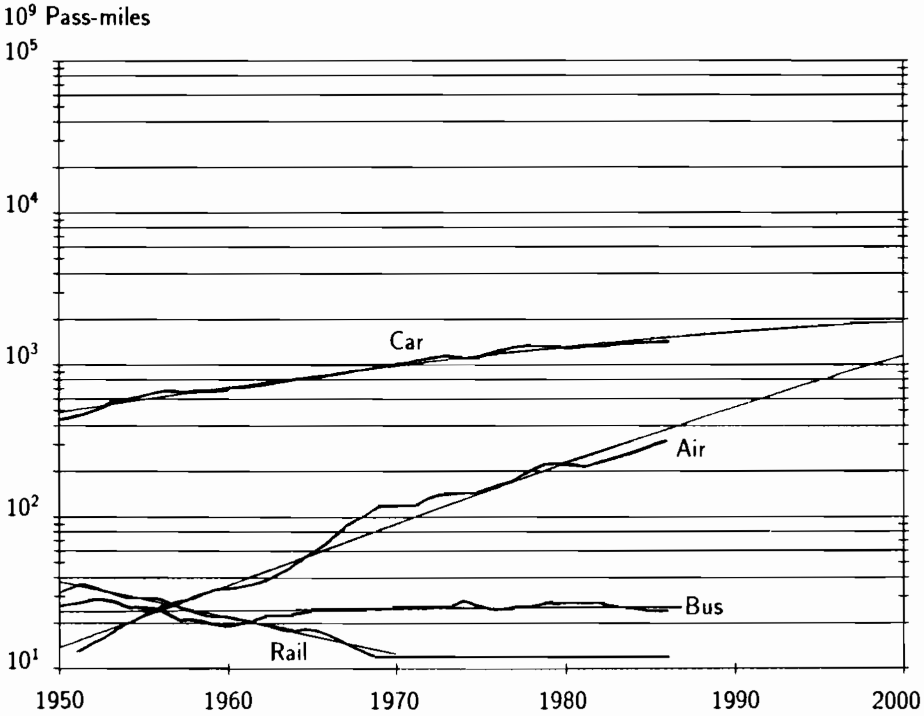


Figure 4.2.1. Volume of intercity passenger traffic in the USA by transport mode, in billion (10^9) passenger miles. (Data source: US DOC, 1977, and 1980 to 1987.)

The analysis is based on an updated* analysis from Nakicenovic, 1986, and portrays a regular competitive pattern in the market shares of four different transport modes for long-distance passenger travel: private cars, public buses, railways, and aircrafts. The analysis shows railways at the end of their technological life-cycle for long-distance passenger transport in the USA. Since around 1970, railways transport less than one percent of the total passenger-miles of intercity traffic. The situation in the USA therefore precedes similar tendencies in other countries by several decades. Bus transport appears to follow closely a long-term decline trend and based on the model forecast should fall below the one percent market share level by the mid-1990s.

The share of private car transportation had reached its maximum market share already in the 1960s, where close to 90 percent of all

---

* This particular analysis provides a good (*ex post*) test of the forecasting capabilities of the competitive evolutionary model used here for the organization of our empirical data base. The estimates as reported in Nakicenovic, 1986, based on data up to 1980 turned out to be excellent forecasts. No market share level of any technology forecasted for the year 1986 deviates more than 4 percent from the actual data.

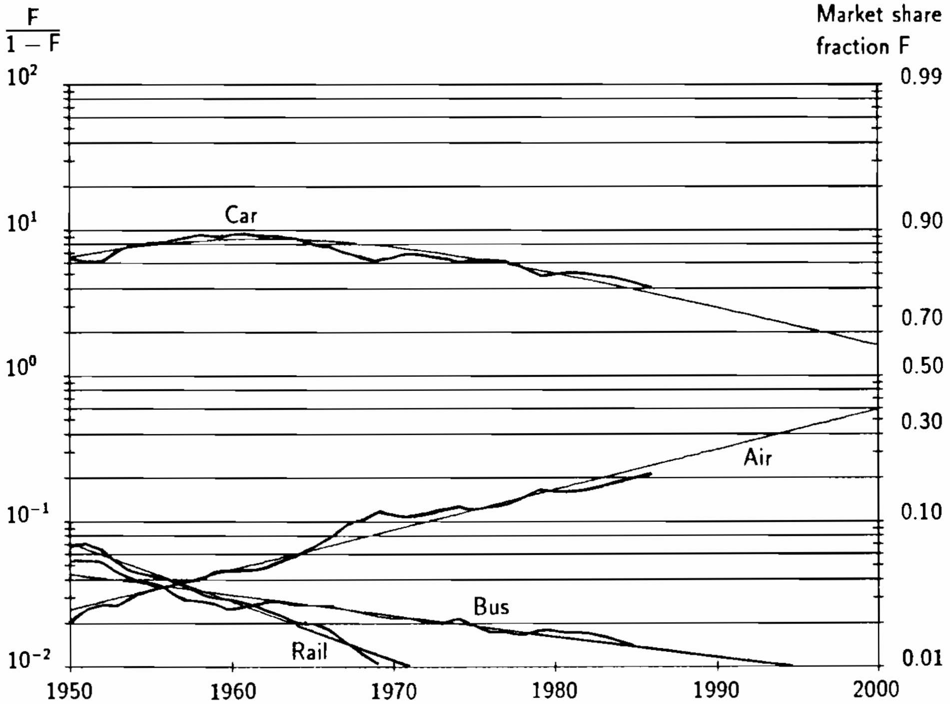


Figure 4.2.2. Modal split in intercity passenger traffic in the USA, measured by fractional shares in passenger-miles travelled. [Source: updated (US DOC, 1987) from Nakicenovic, 1986.]

intercity passenger-miles were accounted for by private cars. Since then, their share has been decreasing due to the rapid expansion of air transport. The share of air transport in total long-distance passenger travel increases with a Δt of around 70 years (symmetrical to the decrease in the market shares of cars). Air travel presently accounts for around 18 percent of intercity traffic and if the long-term growth tendency continues, it will become the preferred mode of long-distance passenger travel in the USA after the year 2000.

The growth in the volume and market share of air traffic does not appear to have been constrained by external forces such as the evolution of fuel costs or by the "striking back" of competitors for long-distance travelling. One can possibly identify, at least in the case of the car industry of the USA, attempts to imitate the successful new competitor, at least *formalisticly*. When the relative market importance of automobiles was at its peak, and started to be challenged by air transport, some cars exhibited secondary design characteristics of aircrafts such as fins very much resembling air foils. Cars were built

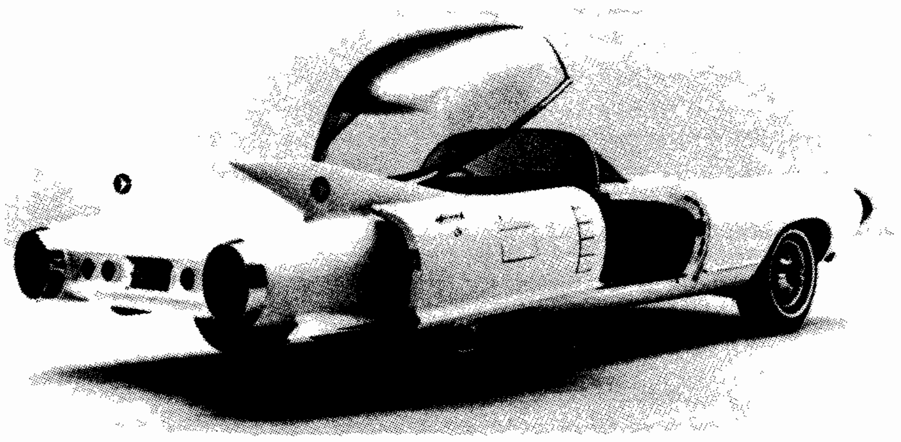


Figure 4.2.3. Response of the automobile industry, when challenged by a new competitor for long-distance travelling in the 1950s. 1959 Cadillac Cyclone, with an automotive heart but designed to look like an aircraft. (Source: General Motors, 1983.)

with an "automotive heart but with an aircraft shell" (Herman and Ausubel, 1988) as shown in Figure 4.2.3.

From a technological history viewpoint, another response by the automobile industry to the rapid growth of air transport can perhaps be put forward, namely there were attempts to incorporate some of the technological developments of aircrafts into automobiles. Car manufactures were, for instance, devoting considerable efforts into introducing the gas turbine into the car industry. The most advanced efforts to develop a car with a gas turbine were probably carried out by Rover in the UK. Similar attempts were made in the USA by Chrysler and General Motors, among others. Prototype models were developed, but never became commercialized. Gas turbines found limited application as a source of motive power for ground vehicles, e.g., for special military (Chrysler tanks) and some early civilian (General Electric and BBC locomotives) equipment. Other more successful examples of aircraft technologies being applied in the car industry include the use of composite materials, turbo charging, compressors, disc brakes, 4-valve engines, and others.

The major driving force behind the structural change in the preferred mode of long-distance travel, and the growth of air travel in particular, appears to be related to the differences in the performance levels of individual transport modes (in particular travel speed) and not so much to the relative transport cost structure. From 1950 to the present, the average operating costs (including fuel) for private cars ranged from 10 to 13 US cents (in constant 1967 US $) per mile (see in particular Section 4.3). If we assume on average two passengers per long-distance car journey, this would mean that the average cost for long-distance car travelling is 5 to 7 US cents (1967) per passenger-mile. Such low levels were reached by air transport only in the mid-1970s, i.e., significantly after the time period when the market share of cars started to decline. For the whole period under consideration, railway transport costs were always significantly lower than car or air transport costs, i.e., consistently below 4 US cents (1967) per passenger-mile. These lower costs did not, however, affect their further decline, both in relative and in absolute (total rail passenger-miles) terms, until the practical disappearance of railways from the market. Our conclusion, that it is not so much economic variables (as reflected in the relative transport cost structure) that appear to influence long-distance modal split decisions, is further corroborated by an analysis of the USSR, which has an entirely different transport and infrastructure policy and transport cost structure.

USSR

An analysis of the long-term changes in the modal split in intercity passenger traffic for the USSR is reported in Figure 4.2.4. Despite some structural differences, we observe a similar dynamic development pattern in the long-distance passenger traffic in the USSR as for the USA. Traditional (slow) transport modes, lose out (logistically) to new competitors such as road and air transport. Road transport is, however, not assured by private cars, but by public buses instead. Yet, the long-term picture resembles (with some lags) the situation in the USA. Inland water transport has virtually disappeared as a long-distance passenger transport mode, railways appear to be on the decline (in terms of relative market share) since the 1930s, and road transport is currently entering saturation, to be replaced in the long-term by air transport.

Figure 4.2.5 compares the intercity passenger modal split between the USA and the USSR. A specific characteristic of the USSR modal split compared to the USA is that it combines a structural evolution of different technological life cycles. Whereas inland navigation for long-distance passenger transport has fallen below the one percent market share level since the middle of the 1970s, railways still account for some 37 percent of all intercity passenger-km of the USSR as opposed to under one percent in the USA. Their displacement process, although lagged by some 50 years (t_0 1972 compared to 1921 in the case of the USA), appears to proceed at a similar rate (Δt

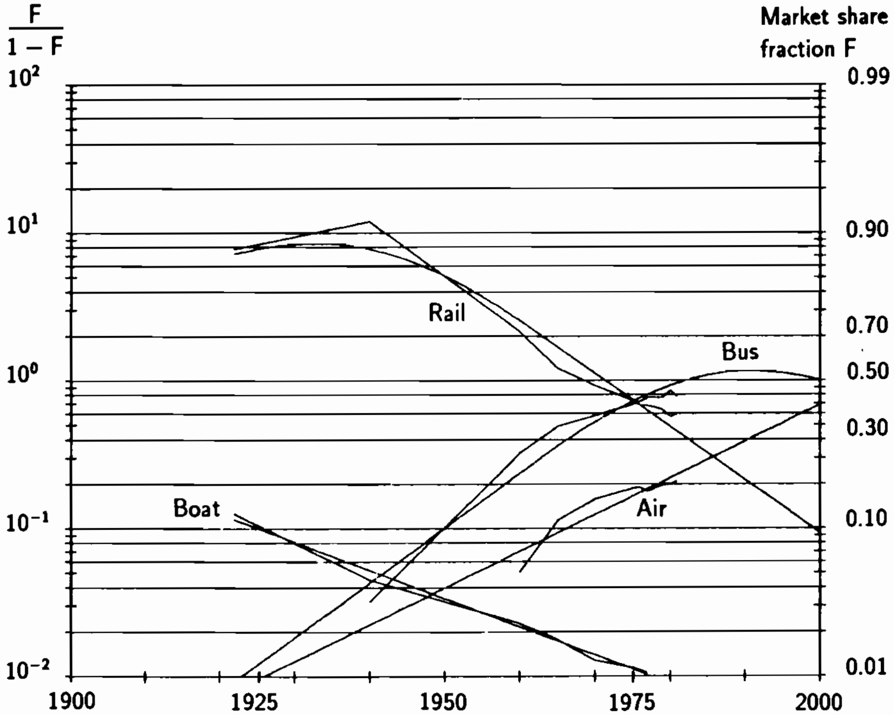


Figure 4.2.4. Modal split in intercity passenger traffic in the USSR, measured in fractional shares of passenger-km travelled. (Data source: Narod. Khoz., 1982 and 1985.)

of 55 years compared to 48 years in the USA). The saturation in the market share of road transport (buses) will seemingly occur some time in the 1990s, i.e., some 30 years after the saturation of the market share of cars in the USA. The most striking similarity however occurs in the area of air transport. Air transport in the USSR presently accounts for some 18.5 percent of all intercity passenger-km, compared to 17.6 percent in the USA. The dynamics of the gains in the market share of air transport in the USSR (t_0 estimated to occur in 2006, Δt 77 years) closely approach the market dynamics of air transport in the USA (estimated t_0 in 2008, and Δt of some 70 years).

This points to very similar *comparative advantages* of transport modes in the long-distance passenger modal split between the two countries, as reflected in their similar rates of change, Δt. This is applicable to negative comparative advantage such as the decline in importance of railways, or positive comparative advantage, as in the case of growing air transport. This similarity in the dynamics of structural change in modal split is important in view of the differences between a market and a centrally planned economy, and especially

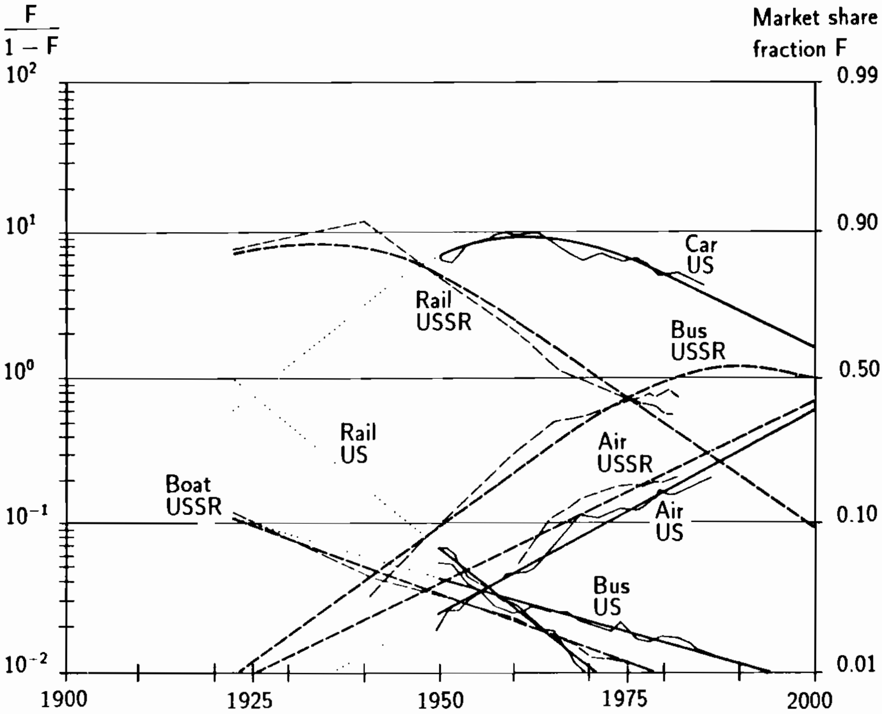


Figure 4.2.5. Modal split in intercity passenger traffic in the USA and the USSR, measured in fractional shares of passenger-km travelled.

with respect to the policies pursued for alternative transport modes in the two countries: Promotion and further railway construction in the case of the USSR, compared to a decrease in network size and the effective disappearance of railways as an intercity passenger transport mode in the case of the USA; road transport by public buses as opposed to private car ownership; a monopoly by the largest airline company in the world, AEROFLOT versus vicious competition between private airlines in the USA, etc.

We therefore assume that the inherent advantages of air transport in altering the human space-time activity framework, i.e., the increased travel range resulting from higher transport speeds by air, is responsible for its gains in market shares in intercity passenger transport. This comparative advantage, resulting from fundamental human time allocation mechanisms, influences individual time budgets and consequently the long-distance transport mode preferences of people, regardless of differences in economic systems.

Therefore, we foresee for the year 2000 and beyond a similar structure for long-distance passenger traffic in the two countries. Railways will have practically disappeared (in the USA) or account for only a few percent of passenger-km travelled (USSR), whereas private (USA) or collective (USSR) road transport and air transport will (with approximately equal shares) become the dominant long-distance transport modes. In the long run, the importance of air transport is expected to increase further at the expense of road transport, if in turn air transportation is not challenged by the appearance of a new (high-speed) long-distance passenger transport mode. In view of the long lead times in the introduction and growth in the market share of new transport technologies identified above, it appears unlikely that the impact of such a new system could be felt before the first decades of the next millennium.

FRG

We now consider a European case – the FRG* An analysis of the long-distance passenger modal split in the FRG as compared to the USA and the USSR is interesting in that we can verify if similar dynamic tendencies can also be observed for a relatively small, and densely populated country, where the distances travelled are not so large and the comparative performance levels of different transport modes (i.e., their speed) should not exert such a decisive influence. As Figure 4.2.6 shows, one can observe a similar dynamic situation in the case of the FRG. The only noteworthy difference is that air transport, due to the shorter distances travelled, accounts for a smaller share in the long-distance passenger traffic and is growing at a much slower rate than in the USA and the USSR. Still, even with a relatively slow growth in the market share, air transport would, based on our model extrapolation, account for more passenger-km travelled than railways by the mid-1990s. If such tendencies materialize, current concerns regarding the development and investments into different long-distance passenger transport infrastructure should possibly be questioned. Based on such a premise, would it not be better if infrastructure development policy were directed at improving airports and the linkages of airports to complementary distribution infrastructures, rather than upgrading the traditional railway network.

Figure 4.2.6 shows private cars as being the preferred mode for long-distance passenger traffic in the FRG, a growth path which was

---

* Data source: *Verkehr in Zahlen*, 1975, and 1985, and Stat. BA., 1987. Long-distance passenger traffic was estimated by assuming that all air passenger-km travelled are long-distance; for railways data for the *Schienenfernverkehr* (i.e., travel range above 50 km) were taken from the above statistics. For bus and private car transport, long-distance passenger-km estimates were derived from DIW, 1982, and 1985. Only data since 1960 have been taken into account, as earlier data exclude Saarland and Berlin and are therefore not directly comparable to later years. In addition, no disaggregation into short- and long-distance private road (cars) transport was available prior to 1960.

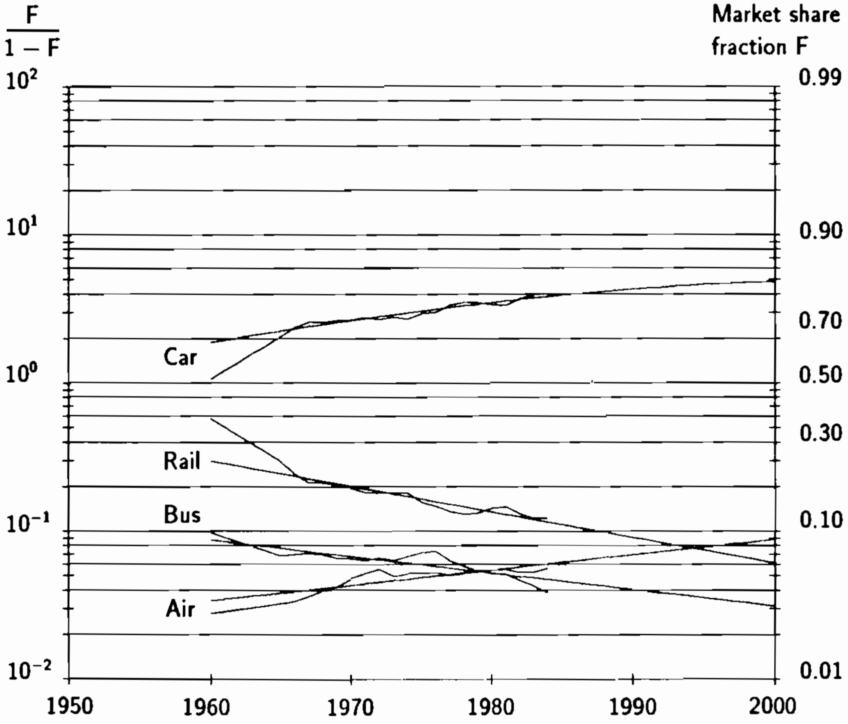


Figure 4.2.6. Modal split in long-distance (intercity) passenger traffic in the FRG, measured in fractional shares of passenger-km travelled.

apparently not noticeably affected by the rapidly increasing gasoline prices in the 1970s. In the medium-term, the share of private cars in long-distance passenger traffic appears to saturate, which could provide an opportunity window for introducing new fast, high throughput ground transportation systems. Their possible success will most likely depend on how effectively they can correspond to individual travel preference choices (i.e., allow for larger distances travelled within a given travel time budget), and how they complement existing high-speed long-distance transport modes, i.e., air transportation. The latter requirement would appear to be of particular importance within a new hierarchy in the long-distance transport chain: air for long distances; new rapid, e.g., rail based, transport systems connected to airports for medium distances; and finally, (private or rented) car transportation for final spatial coverage, complementary to the transport modes at higher hierarchical levels in the transport chain.

The stable dynamic behavior of the long-distance passenger modal split in the FRG, makes it appear unlikely that (interventionist type of) policy measures or drastic changes in the relative transport cost structures could influence the modal split significantly in the short-term. We assume that such changes are possible only by the introduction of new transport systems, with improved performance (speed) and service (e.g., along the line of the railway trains operated by the German Lufthansa) levels, i.e., systems which could alter the space-time framework of human activities over long distances.

France

The evolution of the long-term passenger modal split in France is a special case in our analysis. The excellent historical transport statistics enabled us to reconstruct passenger modal split data* as of the beginning of the 19th century. Here one has to mention the study by Toutain, 1967, who provided much of the original estimates and historical references used to reconstruct the data series. This unique situation will not recur for many years to come in other countries, since long series on (private) road transport are almost entirely lacking.

Unfortunately the long historical series do not allow us to differentiate between long-distance and local passenger traffic. Therefore a comparison of the passenger modal split in France with the figures presented above for the FRG, the USA, and the USSR is limited. Also, as shown below in Figure 4.2.7, the picture that emerges is rather turbulent, even to the extent of approaching the limits of a continuous application to the simple technological substitution model used here for data structuring. The model fails to capture in particular the very rapid introduction of the railways prior to 1850. Recall here that prior to 1830 public railways were practically nonexistent, whereas the model would put the introduction of railways (if the diffusion pattern after 1845 is "backcasted") at a significantly earlier date. Still, Figure 4.2.7 enables us to obtain a concise overview of 160 years of passenger transport in France.

Before the advent of the railways the dominant modes of passenger transport in France were horse carriages and coaches, and barges on inland waterways. The improved travel speed and comfort offered by railways resulted in the rapid displacement of the share in

---

* Data source: Inland water navigation: Estimates by Teisserenc, 1845, [passenger transport by road and (steam) navigation] corrected for road passenger transport (by horse), estimated by Toutain, 1967, (based on contemporary traffic throughput measurements). Rail transportation (incl. TGV data) from Ann., Stat., 1961, and 1973 to 1987, and Mitchell, 1980. Road transport by cars and buses prior to 1965 was obtained from Toutain, 1967, and has been complemented by own estimates for bicycles (especially important before the turn of the century) and motorcycles. For 1965 onwards, IRF statistics (IRF, 1970 to 1985) and national statistics (Ann. Stat., 1972 to 1987) were used as a source for road transport passenger-km. Air transport statistics are from Ann. Stat., 1961, and 1973 to 1987 (referring to all operations, domestic and international). All data were obtained from annual time series, except horse road transport (interpolated from the quintannual data given by Toutain, 1967) and (motorized) road transport prior to 1925, again interpolated from the figures given by Toutain, 1967.

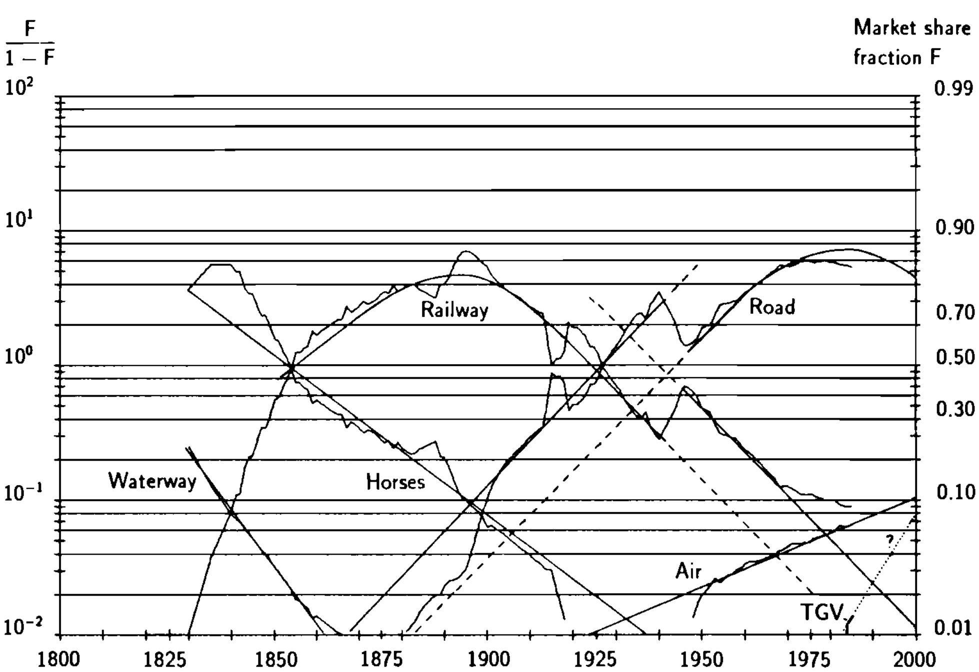


Figure 4.2.7. Passenger (intra- and intercity) modal split in France, fractional shares of passenger-km performed and model estimates. Growth trend for TGV is speculative.

the market of the two previously dominant modes of passenger travel, despite efforts to improve the performance of traditional transport modes, e.g., the introduction of steam power on inland navigation. Inland waterways disappeared as a passenger transport mode by the 1850s, whereas the market share for horse transport declined at a slower rate and finally disappeared with the advent of the automobile around the turn of the century.

The growth in the market share of road transport depicted in Figure 4.2.7 started around 1880, i.e., significantly before the introduction of automobiles. This points to another frequent feature in the introduction of new transport systems: the ground is prepared by previous technologies. In the case of road transport these pre-automobile age technologies were the bicycle and later the motorcycle* As soon as

---

* About 10 percent of all passenger-km in France in 1900 were travelled by bicycles and motorcycles, also for longer, intercity distances. Automobiles were entirely negligible (around 3,000 cars registered) in this early phase of the growth of road transport. The data for bicycles and motorcycles are, of course, only first order magnitude estimates, but are rather on the low than on the high side.

motorcycles, and later automobiles, appeared to expand the competitive niche opened up by the bicycle, road transport started to replace railways along a logistic substitution pattern, proceeding with a Δt of around 60 years. The turbulence in this substitution trajectory due to World War I was elastically absorbed and the process proceeded regularly up to 1939.

However, the effect of World War II and the occupation of a large part of France by German troops was dramatic. This event represents a major political, social, and as Figure 4.2.7 shows, technological discontinuity in the long-term development pattern of France. The disruption appears in fact (contrary to World War I) to be so strong, resulting in a major structural discontinuity and break in a long-term trend, as is represented in the diffusion pattern of (motorized) road transport. During the war years the historical diffusion pattern is reversed, and the previous diffusion level is attained again only after about 15 years. It is as if France went through a period of "hibernation", a major discontinuity and break in long-term societal and technological trends, as manifest* in the long-term substitution between different transport modes. The effects of this discontinuity is reflected in the fact that no continuous technological substitution model application is possible.

Another interesting fact emerges from Figure 4.2.7 – the *diffusion rates* (i.e., the slope of the substitution trajectories depicted in Figure 4.2.7) apparently did not change due to this discontinuity. Thus, the system (i.e., its long-term substitution pattern) "hibernates" for 15 years, in order to reassume the pre-war diffusion pattern with the same diffusion rates (see the parallel trajectories for railways and road vehicles in Figure 4.2.7). According to diffusion theory this is an indication that the relative comparative advantages (defining the speed of diffusion, i.e., Δt) between different transport modes were *not* affected by this major discontinuity. The comparative advantage of road transport as represented in its Δt of around 60 years remained the same in both time periods, before and after the discontinuity, i.e., between 1900 to 1938 and after World War II. This implies that the comparative advantage of road transport (primarily automobiles) responsible for the decline of the railways, was no different before World War II than it was in the period of massive motorization thereafter. From such a perspective, the decline of the railways

---

* In fact the same "hibernation" effect can also be observed in a number of additional societal and technological indicators, including communication behavior (see Section 4.3 below), cultural activities, primary energy balance (replacement of coal by oil showing a similar picture to railways and motorized road transport in Figure 4.2.7) and others.

appears to have been "programmed" from as early as the beginning of this century and is *not* the result of post World War II developments.

A corollary to the above observations is that we might discuss the relative comparative advantage of different passenger transport modes as reflected in their penetration rates (Δt) into the passenger transport market. From such a perspective, of all modes covered in Figure 4.2.7, railways offered the largest comparative advantage in their early phase of development i.e., considerably larger than, for instance, the introduction of the automobile. Railways appeared around 1830 in France and by 1855 some 50 percent of all passenger-km were already transported by railways (i.e., a market share growth rate Δt^* of 25 years) as a result of the large performance differential (transport speed and geographical coverage) railways enjoyed over traditional transport systems. The share of inland water transport decreased with a Δt of around 45 years and that of road transport by horses with a Δt of around 80 years. Whereas inland water passenger-km decreased both in relative and in absolute terms, and by 1860 inland navigation had virtually disappeared as a long-distance passenger transport mode, passenger-km with horse vehicles continued to rise in absolute figures well into the 1880s. This is a further indication of the complementary character of short-distance road transport of horses to railways, a market niche lost by horses only with the appearance of new road vehicles, first the bicycle, and later the motorcycle, buses, and private cars.

Thus, the comparative advantage of railways in their early phase of growth (i.e., prior to 1855) to other transport modes was very high indeed, and higher against inland water navigation (Δt 45 years) than for total (i.e., long- plus short-distance) road transport (Δt 80 years). Only after railroads had effectively eliminated all other long-distance passenger transport modes (river steam boats and stage coaches) did the diffusion pattern slow down (to a Δt of some 80 years). We conclude, that the effective disappearance of long-distance passenger transport by river or by stagecoach must have occurred no later than 1880, based on the dynamics of technological competition identified with our simple model.

Symmetrical to the growth of road transport, railways have entered a declining market share trajectory since the beginning of this

---

* Estimated, as no continuous model application was possible. The reason for this lies most likely in the inadequacy of the data base to allow for separate analysis of long- and short-distance transport modes. The relatively slow replacement of horses is in fact explained in that horses fulfilled an essential complementary role in distributing passengers and goods to and from railway stations. By extrapolating the fast introduction of the railways, one would arrive at the conclusion that by 1880 the horse had already disappeared as a *long-distance* passenger transport mode, which is consistent with the available data on the discontinuation of the last stagecoach services around that time period (see e.g., Voigt, 1965).

century, interrupted only by a shift in its phase-out due to the "hibernation" during World War II and its aftermaths. Since 1970 the rate of decline has, however, slowed down, indicating that railways are apparently phasing out more slowly from their last remaining (below 10 percent) market niche which largely consists of commuting and subsidized transport (e.g., for school children, military personnel, etc.). After its period of spectacular growth and market dominance, road transport now appears to be saturating due to the growth of new competitors. The slower growth rates in air travel in France, compared to the situation in the USA or the USSR, is a direct consequence of the shorter distances travelled and of the resulting smaller comparative advantage (in terms of reduced travel time) compared to other transport modes.

After having discussed the historical developments in the French passenger modal split, what is the outlook for the future? The outlook relies on an important (and in the long-term unrealistic) *ceteris paribus* assumption, i.e., the possible appearance of a new transport technology is not considered. Although there are doubts about whether the TGV is an example of a new system, it could well be that the TGV is the forerunner of new ground based high-speed transport systems such as Maglev's. In Section 3.2 we used the analogy of the *horse* railway between Linz and Budweis as an example of such a transitional system, combining a new infrastructure network with traditional propulsion technology. Under such an hypothesis, we have separated data relative to the passenger-km transported by the TGV in order to provide some insights into the dynamics of the passenger modal split responding to the emergence of a new technological system.

The forecasts of Figure 4.2.7 indicate road transport at the peak of its technological life-cycle (in relative terms; absolute passenger-km by cars is likely to continue to increase in the future). Air transport appears to grow at a steady, yet slow diffusion rate, an indication of the relative short transport distances covered in France, where the comparative advantage of higher speeds of air transport cannot be fully translated into high market share gains. Still, based on our simple model one could expect by the year 2000 that about 10 percent of passenger-km will be transported by aircraft in France, at least for long-distance* operations. As mentioned above, the decline tendency of the railways appears to have slowed down since 1970. In fact, it is a

---

* Here one has to note, however, data problems of the air passenger-km time series available, as they include not only domestic but also international destination passenger-km. Domestic air passenger-km possibly do not account for more than 1 to 2 percent of the total passenger-km (including local and urban traffic) in France. The passenger air transport trajectory from Figure 4.2.7 is thus best interpreted as representing the share of air transport in long-distance (intercity) passenger travel.

frequent observation in the analysis of technological substitution processes, that the last ten percent of market shares are lost at a considerably slower rate than the prime market segments. This may be due to institutional reasons as well as the maintaining of certain (protected) market niches (significant tariff reductions for school children and students, the elderly or people on military service are granted on most European railways). Based on such a premise railways could still account for some 6 percent of passenger-km (compared to 8 percent at present) in France in the year 2000.

Road transport in the longer term appears to be really challenged only by the emergence of new transport modes, such as rapid rail based intercity connections. The success of the French TGV on certain routes (e.g., Paris-Lyon) might be a first sign of the likely market response, once such new advanced high speed ground transportation becomes available. In the absence of a sufficient historical data base to calculate a future scenario, we assume a speculative growth trajectory for the TGV in Figure 4.2.7 to illustrate the impact on the passenger modal split in France. The penetration rate assumed for the scenario is close to the maximum observed over 200 years, i.e., similar to the one prevailing in the introduction phase of railways. This does, of course, not imply that we consider such a growth scenario realistic for the TGV proper. Instead, it is only indicative of the possible maximum penetration rates which could be achieved by a new transport system with a clear cut competitive edge over traditional transport modes, in particular road traffic. If history is a guide, one might expect a continuation of the rapid diffusion of the TGV or similar systems (like Maglev's) up to a level of a few percent market share, and then a transition to a slower substitution trajectory, consistent with observed historic market penetration rates. Even in a speculative high growth scenario, the dominance of the automobile in passenger transportation in France would not be affected noticeably before the turn of the century, pointing to the slow response times in the structural changes in the passenger modal split.

4.2.2. Modal split for goods transport

Goods transport is a market segment quite different from passenger transportation. As we have seen above, the market dynamics of the premium transport market (moving people) are primarily influenced by the inherent performance levels (transport speed) of individual transport modes and appear to be highly price insensitive (inelastic). The situation with goods transport is different in that only for a few, high-value goods (e.g., mail) are similar price insensitive transport mode choices made. In brief, the modal choice for goods transport is a function of availability (access to infrastructure), technological performance (transport speed, geographical coverage, weight limitations, etc.), and finally costs. This complex vector of variables may change over time, especially in response to changes in the performance and price structure of different transport modes. Thus, the long-term dynamics of the modal split for goods ought to be analyzed as a function of price *and* performance rather than as a simple time trend pattern.

France

In view of the above arguments, it should not be surprising that the long-term modal split for goods in France (Figure 4.2.9 below) finally provides an example for which no simple, continuous time dependent technological substitution model application is possible.

The empirical data* indicate a more complex dynamic pattern in the structural change of the goods transport sector in France. Figure 4.2.8 presents the goods transport output by different modes in France since 1800, measured in 10^9 ton-km. The evolution of total goods transport volume is characterized by an exponential growth trend, which is interrupted only by discontinuities such as the war of 1870, World War I, and especially World War II. This long-term exponential growth path of total ton-km transported appears, however, to slow down after the 1970s, pointing to a possible decoupling between economic and goods transportation growth. With regard to the development pattern of different goods transport modes, one can discern two phases in the development of transport output (ton-km) performed. A very rapid introduction phase during which a new transport mode (railways, trucks, pipelines or aircrafts) conquers specialized market niches and in which its comparative advantage is greatest. This rapid initial growth path in ton-km transported is, however, not sustained for long. The growth rates slow down considerably, although still remaining basically exponential for extended time periods. This transition in the long-term growth rates of individual

---

* Data source: Inland water navigation: Data series as reported by Toutain, 1967, and Mitchell, 1980, and updated with Ann. Stat., 1980 to 1987. Data refer to inland (canal and river navigation) only, exports by ship are excluded. Coastal shipping data available in Toutain, 1967, could not be updated and had consequently to be excluded in the analysis. This, however, would not affect noticeably the long-term (declining) trend of the share of water navigation in domestic goods transport. Railway t-km are derived from Mitchell, 1980, (but are consistent with Toutain, 1967) and updated by Ann. Stat., 1980 to 1987. Goods transport by horses and trucks up to 1965 are taken from estimates provided by Toutain, 1967. Quintannual data are (exponentially) interpolated. Since 1965 road transport data are derived from IRF, 1970 to 1985 statistics, complemented by Ann. Stat., 1980 to 1987. Pipeline t-km (oil) are taken from Vivier, 1978, and after 1975 from Ann. Stat. des Transports, 1982 (referring to pipelines longer than 50 km only). Air transport (including exports) data come from Ann. Stat., 1961, and 1977 to 1987.

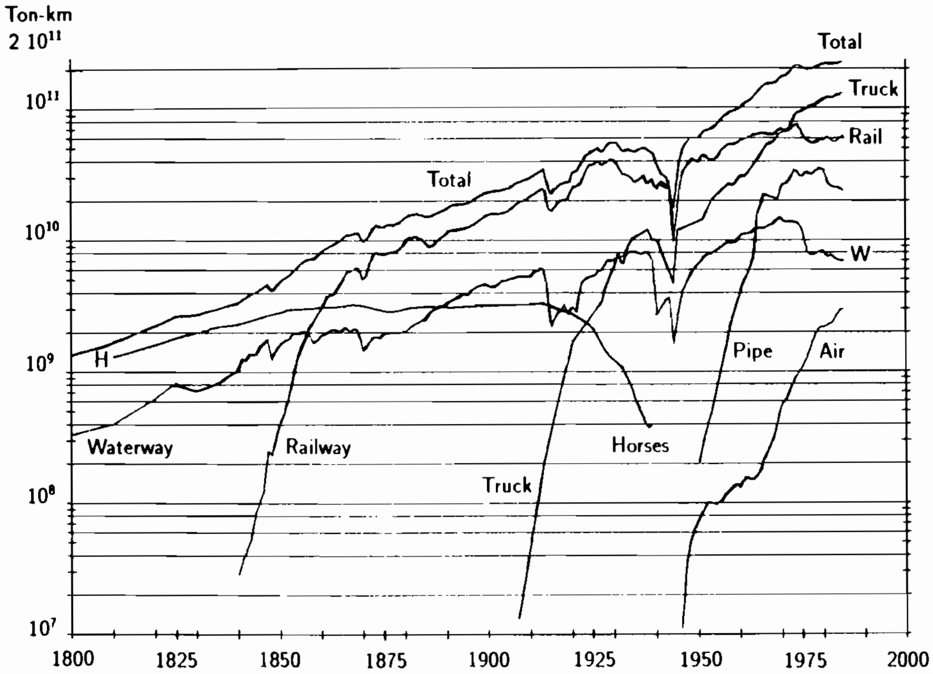


Figure 4.2.8. Goods transport output by different mode and total in France since 1800, in 10^9 ton-km performed.

transport technologies occurred for railways around 1870, for trucks around 1930, and for oil pipelines around 1970. We assume, that this transition in the growth phases of different technologies is triggered by the fact that a new system has, by that date, exhausted its initial growth potential, resulting in the replacement of traditional transport modes from premium market segments.

We see from Figure 4.2.8 that only one mode of goods transport has entirely disappeared from the market: horses, whereas other traditional transport modes, such as inland navigation have continued to transport increasing quantities until fairly recently. Despite the growth in output, the relative share of different transport modes in the total ton-km transported has been changing. This is best illustrated by the role of railways in the goods transport sector. Despite the fact that absolute ton-km transported by railways have increased since the 1920s, their relative market share has been decreasing, as their respective growth rates have increasingly fallen short of the growth rates in total market volume.

Figure 4.2.9 presents an overview of the long-term changes in the relative market shares of different goods transport technologies. The

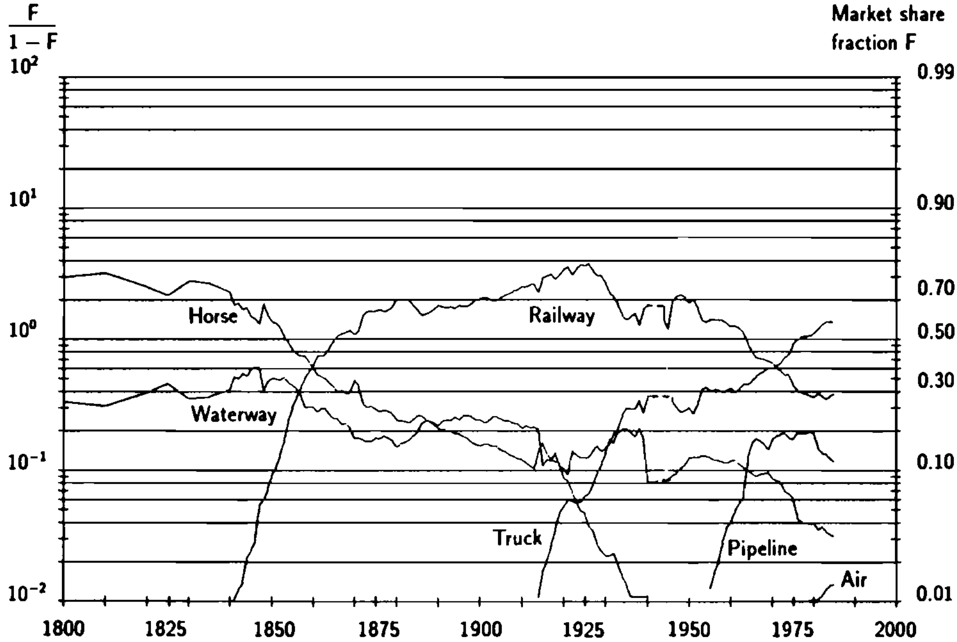


Figure 4.2.9. Modal split in goods transport in France, in fractional shares of ton-km transported.

structural changes in the modal split for goods in France show a complex dynamic pattern, which appears to follow the dynamics of technological substitution only in the introduction phase and again in the last phase-out period. The reason for this was already mentioned above. Diffusion and technological substitution appear to be primarily driven by differences in the comparative advantage of different transport technologies. For low-value goods in particular, the comparative advantage in terms of quality of performance proper (speed, regularity of service, geographical coverage, etc.) does however not constitute the only decision making criterion for modal choice, since (low) transport costs are also decisive. Thus, instead of time, a composite "quality" variable incorporating both price and technological performance (changing over time) ought to be used as an independent variable. Absence of detailed time series do not, however, allow such a comprehensive analysis.

Although no simple continuous technological substitution model can be applied, we nevertheless present in Figure 4.2.9 the long-term modal split for goods in the usual logarithmic $F/(1-F)$ transformation of market shares over time. This is done in order to make logistic growth phases (appearing as linear trends in Figure 4.2.9) clearly discernible from other structural transitions and to provide comparability with technological substitution cases presented earlier.

Even if the complex pattern, which for long periods lacks a significant dynamic aspect (see for instance the share of railways between 1880 to 1930), cannot be formalized by a model, one can nevertheless extract a typical pattern in the introduction of a new mode in the goods transport sector. In the past, new transport modes for goods have typically evolved along a four-phase life-cycle model. Once introduced, their rapid growth quickly displaces existing transport modes, first by carrying higher value goods, i.e., passengers and information, and later on high-value density products. This rapid first phase of market share gains lasted from 1830 to 1870 in the case of the railways, from 1910 to the end of the 1930s for trucks, and from 1955 to 1965 for oil pipelines. The performance advantage of a new transport mode appears to play a decisive role in rapid initial market penetration. Figure 4.2.9 suggests, that this initial "premium" market niche was highest for railways, with their market share increasing from one to nearly 50 percent in less than 20 years. Truck transportation also rapidly filled its first market niche, but, with around 25 percent market share, it was already smaller. The specialized market segment for pipelines never exceeded 15 percent of all ton-km transported.

Once the initial niche has been conquered, growth proceeds at a significantly slower pace with additional shares being gained from the traditional, slower transport modes, primarily via tariff competition, since transport costs decrease along the learning curve. This second phase is completed once a system reaches maturity and complete market dominance. In France, this second phase lasted from about 1870 to the 1920s for railways, when over 70 percent of all ton-km in France were transported by rail. For truck transport this second phase appears to be continuing.

The third phase of the four stage life-cycle model consists of a relatively long saturation period, during which a transport mode loses market shares but maintains a dominant position. For the railways in France, this third phase lasted until the 1950s and 1960s, when their respective market share fell below the 50 percent barrier for the first time. The last phase is the period of fall and decline under vicious competition from newer transport modes. This has been the situation of railways since the 1950s. Railways are retreating progressively into the lowest value market niches of the goods transport sector by transporting bulk, low-value raw materials, like gravel, scrap or coal, where low transportation costs are a more important criteria than quality of service or speed.

Figure 4.2.9 allows us to summarize the long-term history in the (domestic) goods transport modal split in France. At the beginning of the 19th century approximately one fourth of the ton-km carried were shipped by inland navigation and three fourths by (horse) road transport. As a result of canal development, the share of inland navigation in ton-km transported increased to an all-time peak market share value of 38 percent by 1847. The subsequent rapid increase in the share of railways in the middle of the 19th century is an indication of their considerable comparative advantage which affected canals and road (horse) transport in a similar way. As of 1860 the gain in the market share for railways proceeded at a significantly slower rate. These later market share gains occurred primarily at the expense of road transport, with inland water navigation keeping a relatively stable market share around the 20 percent level between 1880 and 1940. From a long-term perspective, inland water transport appears to be the least dynamic (sensitive) transport carrier, losing only gradually and slowly its share of the market to other transport modes. This would be an indication of the fact that inland navigation retained its comparative advantage (low transport costs) in the lower value goods (bulk commodities) market segment, where transport speed is not so decisive.

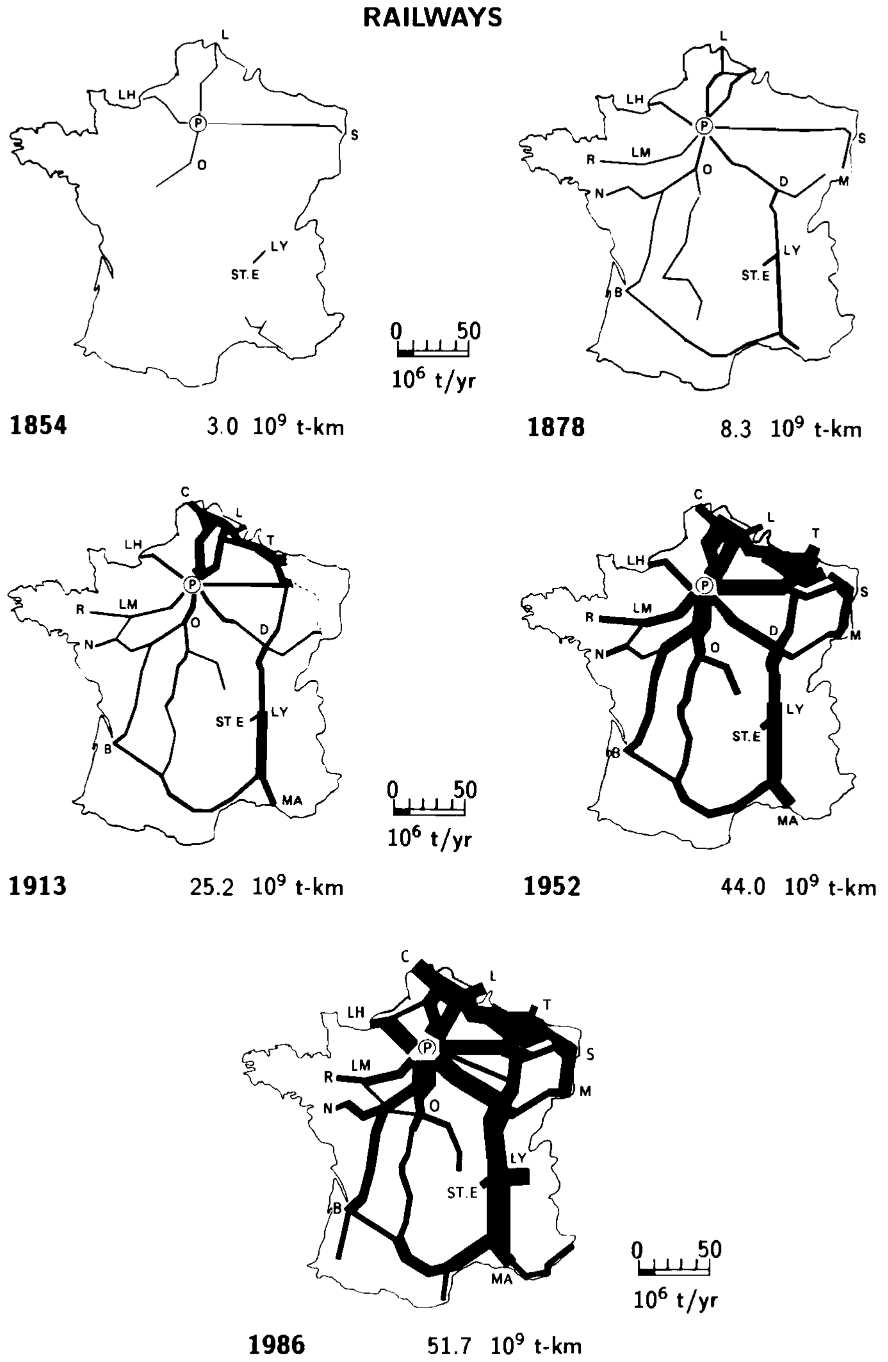
The decline and end of horse road transport due to the rapid diffusion of trucks between 1910 to 1940 has been discussed already in Section 3.3. After the completion of the replacement of the horse in road transport, the market share gains of trucks continued to proceed at a slower rate than during the early introduction phase. This was primarily at the expense of railroads, which after having reached their peak in the market share in the goods transport sector (79 percent in 1926), declined. Also noteworthy is the rapid initial growth (1950 to 1965) of the market share of a specially dedicated transport infrastructure: oil pipelines. Their share in the total ton-km transported reached its maximum in the 1970s with around 15 percent, in order to decline thereafter.

A most likely near- to medium-term outlook for the future goods transport in France would suggest a continuation of the structural shifts apparent since World War II. Such a scenario would imply that road transport could eventually reach a similar dominant position in the transport sector as railways enjoyed in the 1920s, accounting for close to 80 percent of all ton-km transported. Railways and inland water navigation would, by continuing their past trends, decrease further in importance and the relative share of oil pipelines is also not likely to increase again. At present no real technical alternatives, resulting in a similar "breaking" of the market dominance of truck transportation, as was the case with railways some 50 years ago, are apparent. Further gains in the market share of road transport appear likely in the transition of manufacturing in the direction of higher value goods and smaller batch sizes under "just in time" and inventory minimization strategies. To meet these requirements, road and air transport appear to have a clear competitive edge as regards flexibility and speed of delivery as opposed to traditional transport modes. Further increases in market shares and also absolute volumes of truck transport* are likely to trigger further policy measures to reduce negative externalities associated with growing truck traffic. Despite the high comparative advantage of air transport in the goods transport sector, it does not appear likely (in view of the relatively short transport distances in France) that in tonnage terms, air freight would transcend the role of a marginal contributer to the domestic goods modal split. Recall here, that in the statistics presented in Figures 4.2.8 and 4.2.9, exports are included in air ton-km. Thus, the available data overemphasize the importance of air transport in domestic ton-km. The situation with respect to the share of air transport in the *value* of goods transported will however be drastically different as shown below.

To finalize the discussion of the long-term trends in goods transport in France, we illustrate the spatial representation of the growth in the volume of goods transported as shown in Figure 4.2.10. Here the *intensification* in the usage of different transport infrastructures, railways and roads, based on the annual tonnage throughput at various segments of the infrastructure network, is shown. The original data (Renouard, 1960, SNCF, 1987, and SETRA, 1988) have been replotted on a common scale, i.e., the thickness of the lines in Figure 4.2.10 is proportional to the tonnage transported. We also present the total national goods traffic flux in 10^9 ton-km for the railway and road network for the years shown in Figure 4.2.10. As regards railways (Figure 4.2.10A), we observe two phases of growth. The first one being characterized primarily by an extension of the network (between 1854 and 1878), whereas the second phase is characterized by an increasing intensification of the use of the existing railway network. Note in particular, the intensification in tonnage transported between 1913 and 1952 and the only modest increases since then, as well as the high concentration of tonnage transported in the industrial North-East of France, the center of the coal and steel industry.

---

* We discuss below the apparent "decoupling" of goods transport intensity and economic growth. If this decoupling tendency continues, total t-km transported, which increased with a rate of 3.9 percent per year between 1950 to 1985, might considerably slow down, depending on overall GNP growth. However, even under only a gradual growth in total market volume the ton-km transported by trucks is most likely to increase further also in absolute terms.



1854 3.0 10^9 t-km

1878 8.3 10^9 t-km

1913 25.2 10^9 t-km

1952 44.0 10^9 t-km

1986 51.7 10^9 t-km

Figure 4.2.10A. Goods tonnage (tons/year) transported over various segments of the French railway network from 1854 to 1986. Thickness of lines is proportional to annual tonnage transported.

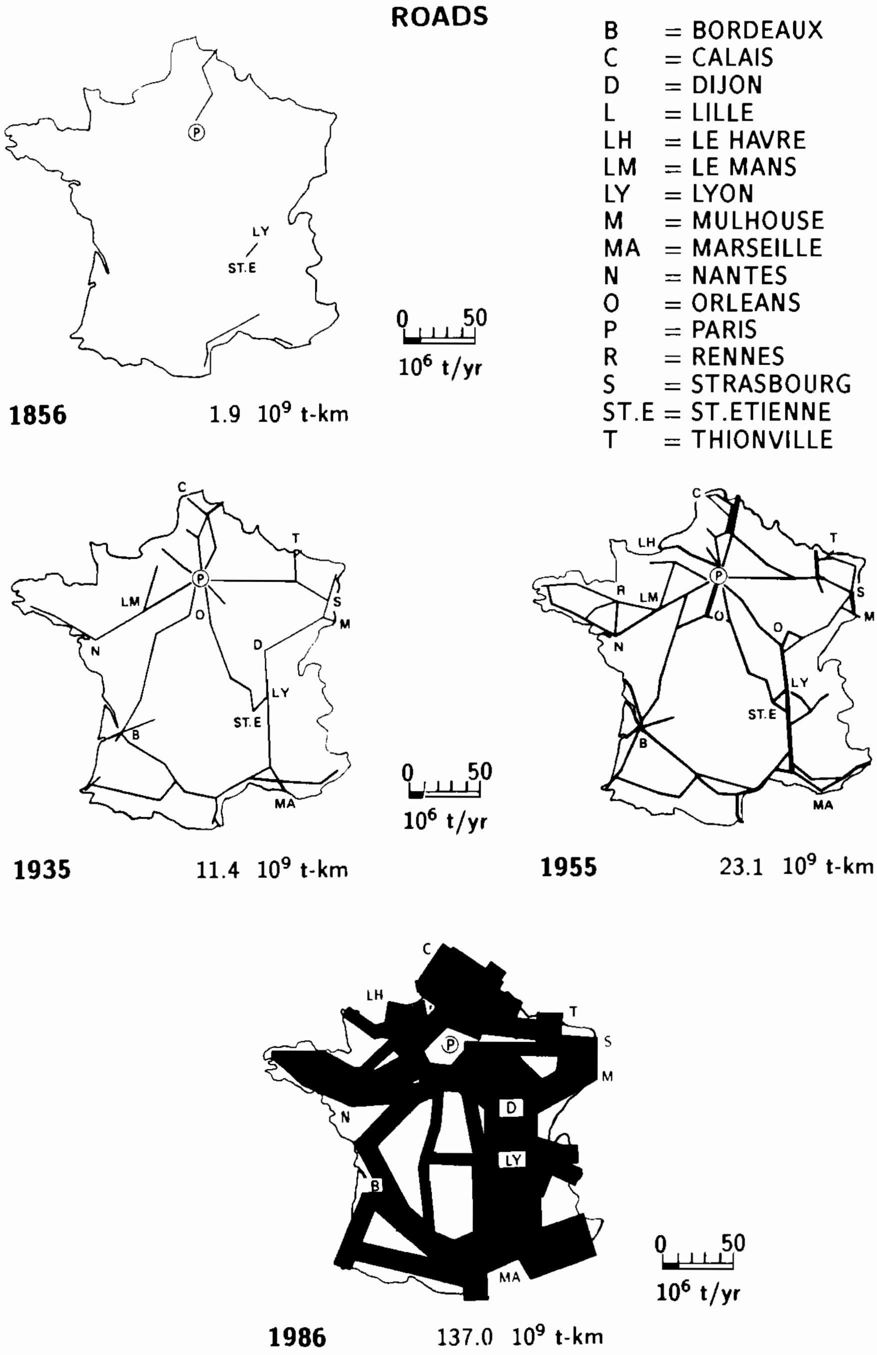


Figure 4.2.10B. Goods tonnage (tons/year) transported over various segments of the French road network from 1856 to 1986. Thickness of lines is proportional to annual tonnage transported.

The dominance of railways in the goods transport sector from the 1850s to the 1950s becomes particularly visible in the comparison to the throughput (tonnage transported) on roads (Figure 4.2.10B). Since that time however, the throughput at various segments of the road infrastructure has increased dramatically: on average by a factor of 6 (measuring ton-km). Despite the fact that in Figure 4.2.10B the 1986 road transport intensity is only an estimate* and may only to a limited degree be comparable to earlier years, it nevertheless illustrates the phase transition in the spatial pattern of the goods throughput between different transport infrastructures in France: from railways to roads. Also visible in the comparison between railways and roads is the increasing spatial division of labor, leading to a polycentric structure, compared to the concentration in the industrial North of France some 70 years ago. Finally, the emergence of a new economic corridor stretching from Paris over Dijon and Lyon to Marseille, as reflected in the intensity of goods throughput both on roads and on railways, is clearly illustrated in Figure 4.2.10.

FRG

Figure 4.2.11 complements the picture for France discussed above, with the modal split for goods for the FRG, in terms of ton-km transported by inland navigation (i.e., excluding sea transport), railways and trucks. The oil pipeline infrastructure, accounting presently for less than 4 percent of total ton-km transported in the FRG has been omitted in Figure 4.2.11, due to a lack of long time series. By means of a continuation of past trends, a similar structural change as identified in France becomes apparent in the modal split for goods in the FRG. Around the year 2000 nearly three-quarters of all domestic ton-km in the FRG would be transported by trucks, with the remaining share accounted for equally by inland navigation and railways.

Air transport does not appear in Figure 4.2.11 at all, as only some 0.1 percent of all ton-km are transported by aircraft. However, present tonnage oriented transport statistics are, to some degree, misleading, as regards the importance of air transport, especially for an open economy such as the FRG. Instead, we consider that another more appropriate indicator would be the value-km (DM-km) transported by different transport modes. Contrary to the passenger modal split, an analysis of goods transport should consider the different value of goods transported (i.e., the heterogeneity of the products

---

* Contrary to earlier years, only vehicle throughput statistics for national roads and highways are available (SETRA, 1988). We have assumed that the 1986 tonnage transported by trucks is distributed proportionally to the average vehicle frequency on the various segments of the road network.

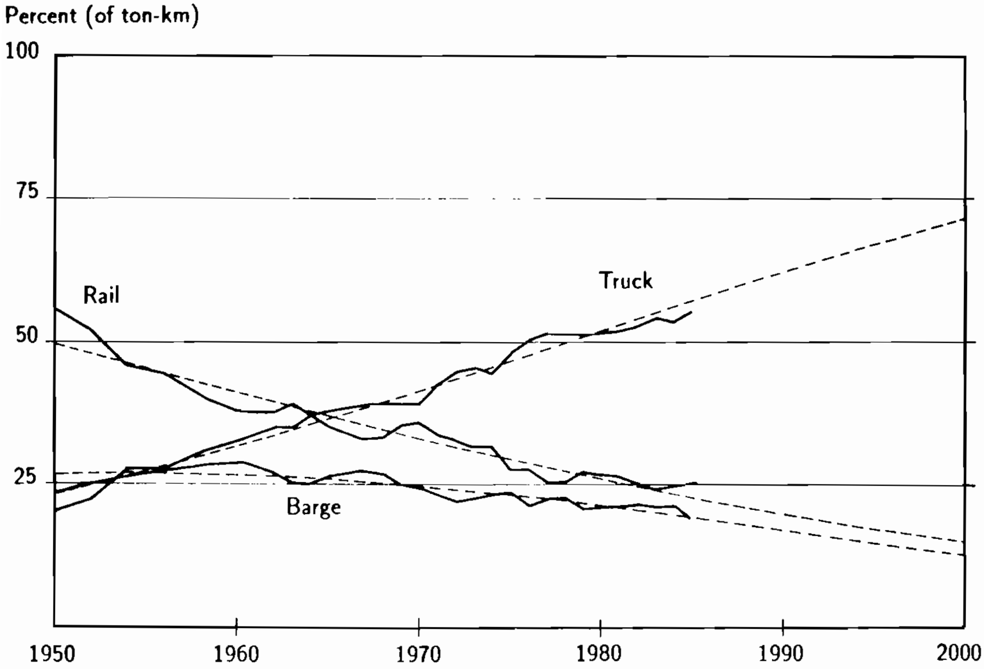


Figure 4.2.11. Modal split in domestic goods transport (measured by ton-km transported) for the FRG (excluding oil pipelines), percentage of market share and estimates based on logistic substitution model.

transported) by different modes. This has been incorporated* into official transport statistics only recently. Some short-term data (since 1969) on the modal split of imported goods in value terms are available, which are further analyzed in Section 4.3. These short-term data (Stat. BA. Fachserie 7, 1979, and 1986, Stat. BA. (WiSta 7/1987), 1987, and Stat. BA., 1987) nevertheless allow interesting insights (Figure 4.2.12) into the relative importance of different transport modes in the *value* of goods exchanged.

Even if the available data only span 18 years and only refer to imports and not to exports (data on the value of domestic transport volume are even entirely lacking), can one see that air transport becomes important, when considering the *value* of the goods transported. Whereas in 1969 some 6 percent of the value of imports was arriving by aircraft, this percentage had increased to over 10 percent by 1986. Sea transport is also an important transport mode for

---

* Export statistics taking into consideration the transport mode and the value of goods transported are compiled from 1987 onwards. Import statistics are available since 1969.

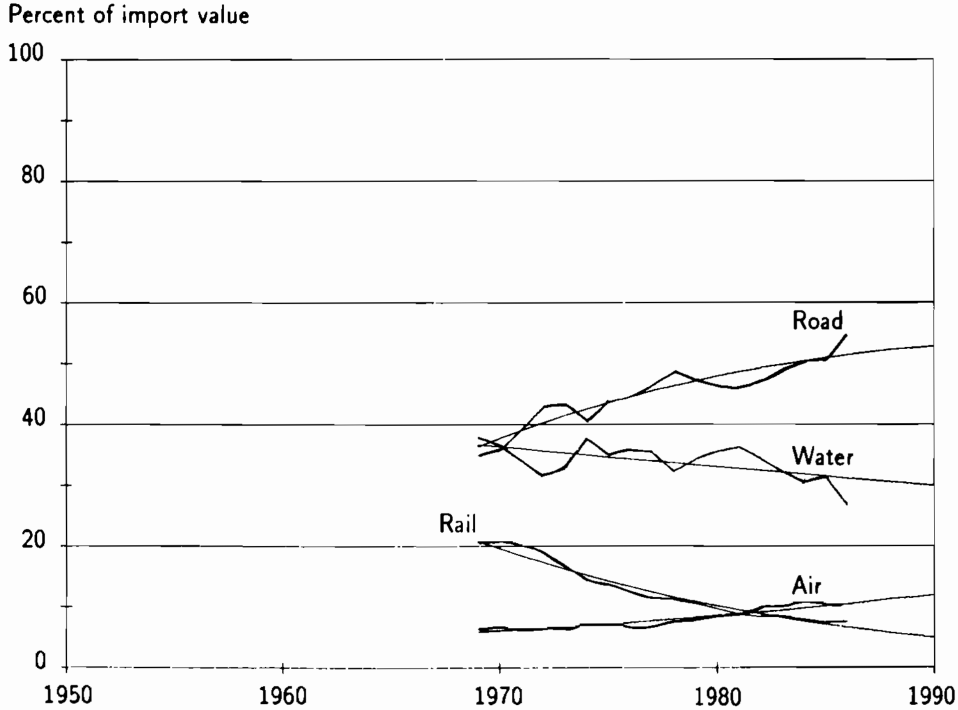


Figure 4.2.12. Share of different transport modes in import value (*Generalhandel*) of the FRG and estimates based on logistic substitution model.

imports to the FRG, but its share in the import value has decreased from 38 percent in 1969 to 27 percent in 1986. The share of road transport (trucks) increased on the other hand from 35 to over 55 percent over the same time period. Most interesting is, however, the situation with respect to railways. Their share in the import value decreased from 21 percent to about 7.5 percent between 1969 and 1986. Thus, at present a higher DM-value of goods is imported to the FRG by air than by rail. Based on our model representation of this structural shift process, the rate of decline of the railways is estimated to proceed with a Δt of 58 years. This compares with a logistic decline rate of railways in domestic ton-km transported (Figure 4.2.11 above) with a Δt of 116 years. This means that the market share loss of railways measured by the *value* of the goods transported (imported) to the FRG proceeded *twice as fast* as their market share losses in the ton-km transported. This demonstrates clearly the structural dilemma of railways in the goods transport sector: while losing tonnage market shares, they lose out even more on the share of the *goods value* transported. At present, it is by no means clear which (drastic) institutional, organizational as well as technological measures (high-value goods normally represent small batch sizes and require high turnaround times, i.e., transport speeds) could be introduced, in order to halt the decline of the railways in the goods transport sector. Otherwise, railways may end up with a similar market niche as inland navigation. While still transporting say around 10 percent of the total ton-km, their market niche would be confined to extremely low-value goods, primarily basic commodities transported in large batch sizes, such as coal, scrap, gravel or wood trains.

Sweden

As a conclusion to our discussion on the modal split of goods transport let us consider the case of Sweden as an example of a small open economy. Figure 4.2.13 presents the share in the *export value* of manufactured goods transported by different transport modes from Sweden to their final destination. As in Figure 4.2.12 above, the share of different transport modes in the value of exported manufactured goods from Sweden is presented in linear form. Actual data and estimates resulting from a logistic substitution model are given, based on data for the period 1965 to 1984. The trend revealed is consistent with the picture emerging from Figure 4.2.12 in the case of the FRG. Railways, and in the medium-term, also sea transport will account for decreasing shares in the value of exported goods in Sweden. Truck and air transport are on the increase, and based on the model forecasts one may assume that by the year 2000 truck and air could account for nearly equal shares in the goods exported from Sweden, measured by the value of the exports. For Sweden, there appears to be a clear tendency for exports to have a higher value content than imports (Snikars, 1987). Thus, exports are increasingly likely to leave Sweden by air or by truck, and only the lowest value goods will be shipped by railway and ship. Thus, our conclusion with respect to the importance of high-value market niches (passengers or high-value goods) in the long-term dynamics of the transport modal split is further corroborated.

The main driving forces for the changing structure of the modal split for goods may, therefore, be summarized as follows: increasing value density of tangible goods and decreasing material intensity tend to generate a higher total value in the transport sector. Higher quality service, smaller batch sizes, and faster and reliable deliveries are required for the transportation of high-value density goods. Despite higher ton-km tariffs, the total transport costs incurred are lower due to quicker turn-around times that allow for a significant reduction in inventories and working capital that would otherwise be tied up in

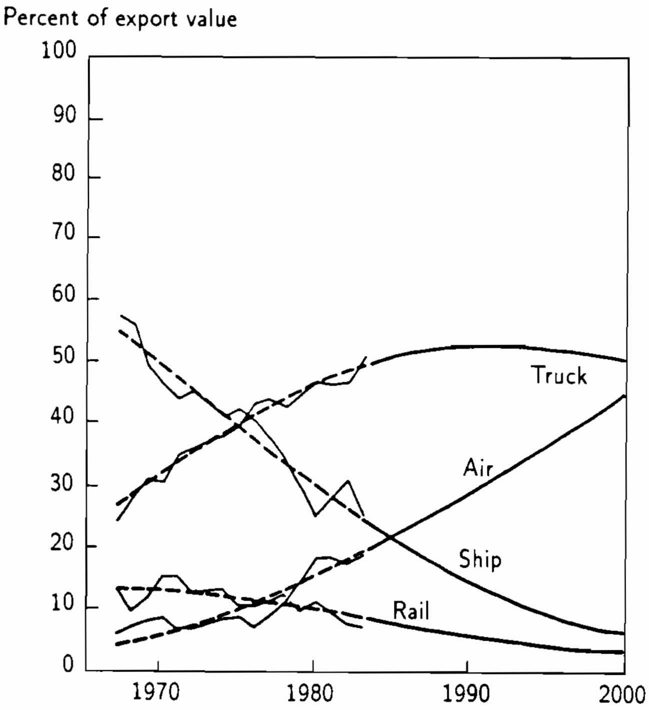


Figure 4.2.13. Share of different transport modes in the export value of manufactured goods in Sweden and estimates based on logistic substitution model. (Source: Snikars, 1987.)

goods waiting to be shipped. These tendencies favor truck transportation for domestic deliveries and increasingly air transport over larger distances, especially for exports and imports.

At this point we conclude the empirical description of the long-term trends in the structural shifts and substitution of different transport modes. During much of our phenomenological discussion, we have already made conjectures about the principal driving forces responsible for the long-term structural change in the transport sector. These are discussed in more detail in the Section 4.3. We consider technological performance as represented in quality (primarily transport speed for the whole transport chain) and flexibility of service, to be the principal driving force. The comparative performance advantage of a new transport system first becomes translated into market share gains in the premium (high-value) market niches of the transport sector, passenger transport, and transportation of high-value goods (including information).

Older technologies, failing to comply with the evolving requirements of the transport system as regards performance (speed), quality, and flexibility of service tend to lose these high-value market niches first, in order to become confined in the long-term to transporting only relatively low-value primary materials and commodities. This "last resort" market niche appears, however, not to be a viable survival strategy in the long-term. With the increasing tendency towards dematerialization in the advanced economies and a shift to more information and software content and thus high-value goods, this last, low-value market niche may eventually not be sufficient to allow the long-term survival of traditional transport carriers and infrastructures.

4.3. Driving Forces and the Economics of Modal Split Changes

Having established empirical evidence with respect to the life cycle of individual transport infrastructures and the long-term structural changes in the transport sector, we formulate below some tentative propositions concerning the driving forces behind these evolutionary processes.

We emphasize that we consider as a long-term driving force the relative *comparative advantage* of different transport systems, seen as a *complex vector* of technological, social, institutional, economic and psychological factors, rather than as a single variable. With this *caveat* in mind, we propose some tentative clusters of driving force variables that have to be considered as intrinsically interwinded and interdependent, and which cannot be reduced to a one-dimensional view of being the one and only (principal) driving force in the long-term development of the transport system.

The comparative advantage of individual transport technologies and infrastructures discussed originates from three areas: First, technological performance (in particular speed and range). Second, the importance of high-value market niches in the initial development and subsequent growth of infrastructures, as well as the learning curve phenomenon (decrease in real term transport costs) that allows new transport systems to enter competition successively in other than high-value market niches, are emphasized. Third, with regard to future development, we argue that the transition of developed economies toward dematerialization and higher information and software content in products, will have important repercussions on the transport system. Such developments would increase the requirements for fast, reliable transport modes with rapid turnaround times, as a result of inventory minimization and "just in time" production strategies. It appears likely that the long-term goods transport intensity per unit of economic activity of advanced economies may not grow any further, in fact it may even decline. Such developments result from increases in the value generated per unit of physical output. Continued economic growth and increasing personal income will most likely have demand stimulating effects for passenger travel, in particular long-distance leisure travel. Consequently in the near- to medium-term we do not foresee any "decoupling" or even decrease in the passenger transport demand from economic growth.

Technical performance

Figure 4.3.1 presents speed trends in transportation, adapted from an article by Samaras, 1962. Transport speeds, from the horse (represented by the pony express for urgent mail delivery in the American West) to trains, to automobiles and finally to aircraft have increased by over two orders of magnitude over the last 200 years. We have discussed in Section 3.2 above, that the higher transport speeds of railway transport (in addition to denser spatial coverage and higher transport capacity) was without doubt *the* decisive performance criterion, explaining the success of the railways at the expense of horse carriages and inland water navigation (canals). This speed advantage of the railways was, however, not accepted without challenge. A number of medical studies of the early 19th century claimed that any transport speed exceeding about 20 km/hr would be detrimental, even lethal, to human health (see e.g., Voigt, 1965).

As usual, cartoons provide an excellent account of contemporary concerns of possible negative effects from a new technology, as illustrated in Figure 4.3.2. Despite such early scepticism and social resistance, the impact of railways in altering the space-time framework of human activities and industrial production, settlement patterns, etc. was decisive. As a result, by 1900 railways had achieved a dominance in the transport sector in all industrialized countries which was considerably larger than the present dominance of the automobile.

The speed advantage of the first automobiles over horse carriages does not require further discussion. Although the average transport speed of the automobile did not increase along a similar dramatic* trajectory as suggested by Samaras, 1962 (referring apparently to record speeds), its independence in usage and nearly complete

---

* Currently, the average car transport speed in the USA does not exceed 50 to 60 km/hr (including urban traffic).

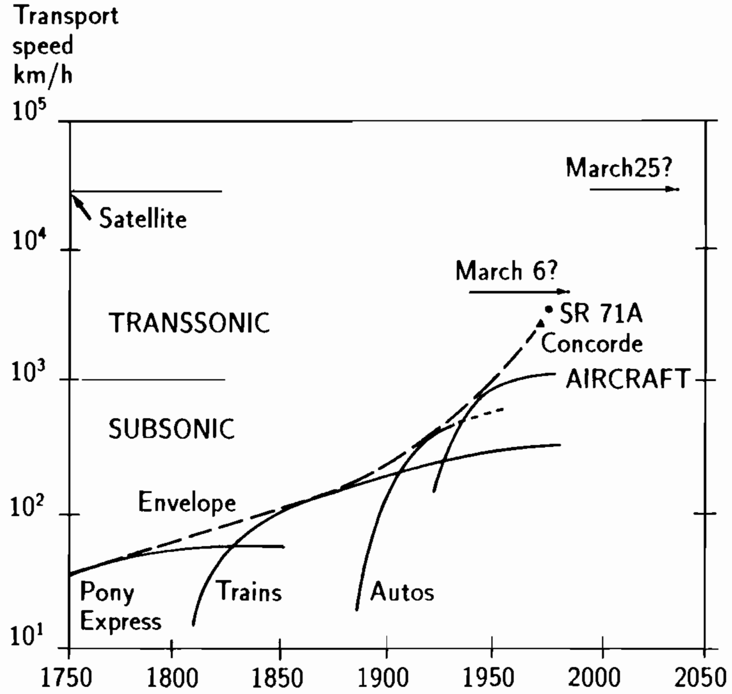


Figure 4.3.1. Transport speed trends, in km/hr. (Source: adapted from Samaras, 1962.)

geographical coverage facilitated a considerable reduction in travel time if we consider the *whole* (i.e., door-to-door) transport chain. Once the automobile had replaced horses as the complementary transport mode to and from railway stations, it also progressively became the preferred mode for long-distance travel and started to compete with railways. In the goods transport sector trucks offered an additional noticeable advantage over railways: smaller batch sizes. Voigt, 1965, argues that this was an important stimulus for the early development of truck transport in the USA, as many companies tried to minimize their inventory (i.e., capital) costs during the Great Depression. Instead of ordering large batch sizes (railway wagons) of primary materials and intermediate goods, smaller, but more frequent deliveries by trucks proved advantageous under extreme competitive pressures. The further growth of the automobile as a form of long-distance transport was also closely linked to the widespread diffusion of private car ownership, the spread of suburban settlement patterns, and the upgrading of road infrastructure, e.g., in the form of *Reichsautobahnen* or interstate highways, allowing for higher transport speeds in intercity traffic. It was concluded in Section 3.3 above, that

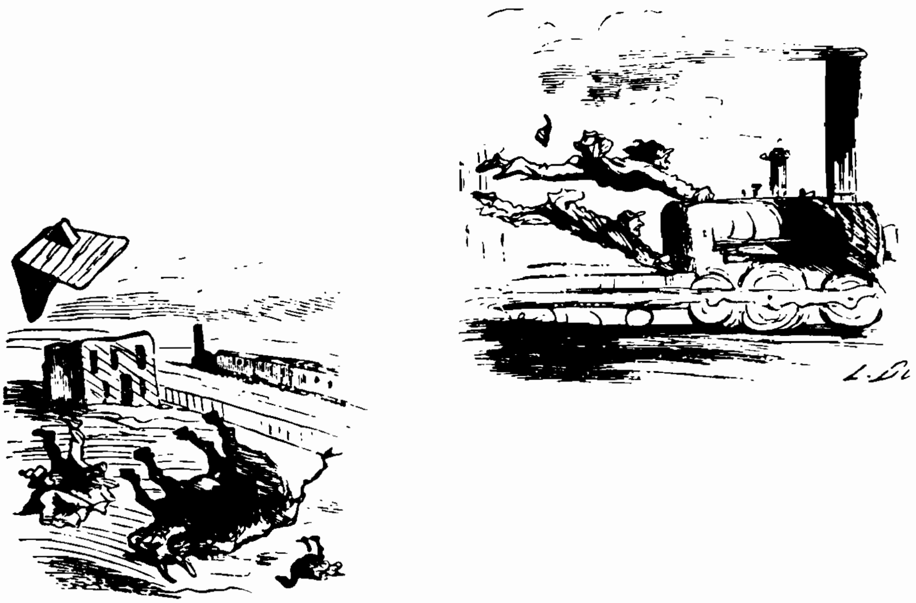


Figure 4.3.2. Impact of a 60 km/hr fast train on the environment and train operators as seen in cartoons from 1862. (Source: Roszak and Schmidt, 1986.)

the relationship between automobile diffusion and the development of road infrastructure evolved inversely (i.e., the growth of the surfaced road infrastructure significantly preceded the diffusion of the automobile). The upgrading of long-distance road connections was however an important factor in the comparative door-to-door travel time advantage of the automobile over railways for medium-range distances (up to a few hundred kilometers), and can be considered as a decisive factor in the long-term substitution pattern in intercity transportation.

In Section 4.2 above, we showed that the automobile started to lose market shares in intercity passenger traffic in the USA because of the growth of a new transport mode, aircrafts, which brought about a quantum leap in the reduction of travel time over large distances. Despite considerably higher transport costs initially, aircraft immediately started to replace automobiles for long-distance transport in the USA. It is therefore natural to assume that the faster transport speed of aircraft was the decisive performance criterion responsible for this substitution. This statement is further corroborated by similar long-term tendencies in the USSR and even in Europe. In Figure 4.3.1

above, the increase in the average transport speeds of commercial aircrafts (using data from Martino, 1983) have been plotted, instead of the original aircraft record speed performance curve from Samaras, 1962. The average travel speed of aircraft has approached the sound barrier, a difficult hurdle to overcome for passenger transport, at least for the present technological generation of passenger aircrafts. The transsonic speeds realized e.g., by the Concorde or military aircrafts such as the Lokheed SR 71A, or the X30 (Mach 6), announced for the mid-1990s appear, however, to be consistent with the envelope of transport speeds as shown in Figure 4.3.1. Under such an assumption, two successive new generations of aircrafts – methane fueled up to around Mach 6, and hydrogen fueled up to Mach 25 – could push the feasibility frontier of transport speeds to a level at which all destinations on the globe could be reached within one hour. Whether such a scenario is realistic cannot be resolved within the present context. Faster transport speed was, and continues to be the most important comparative advantage air travel has over traditional transport modes, especially for premium market segments such as passenger travel and transportation of high-value and perishable goods.

Without taking a monocausality viewpoint, we conclude nevertheless that faster transport speed is to be considered as an important driving force for the long-term competitive pattern outlined in the empirical part of this study. Faster transport speeds imply – under a constant travel time budget constraint – that spatial distances are "compressed" and that a larger (territorial) range can be covered by an individual. This reduction of travel time results in a "space-time convergence" (Janelle, 1969) between locations. For instance the travel time by stage coach between Edinburgh and London improved from some 20,000 minutes* (two weeks) in 1660 to some 2,000 minutes by 1800 (mostly due to improved turnpike roads). The introduction of railways reduced this travel time to less than 800 minutes by 1850 and with further speed improvements to less than 400 minutes at the present time. The travel time between the two cities by aircraft is less than 200 minutes. Thus, the reduction in travel time and temporal distance is a factor of 100 over a time period of some 300 years, i.e., the *effective* space separating the two cities for human activities (measured by travel time) has been "converging" dramatically.

A corollary of the space-time convergence is that the spatial range (distance) that can be covered by an individual per unit of time is increasing. Figure 4.3.3 illustrates this for France, by plotting the average daily distance travelled per capita by transport mode, and the

---

* All data from Janelle, 1969, who also provides a graphic representation of this space-time convergence.

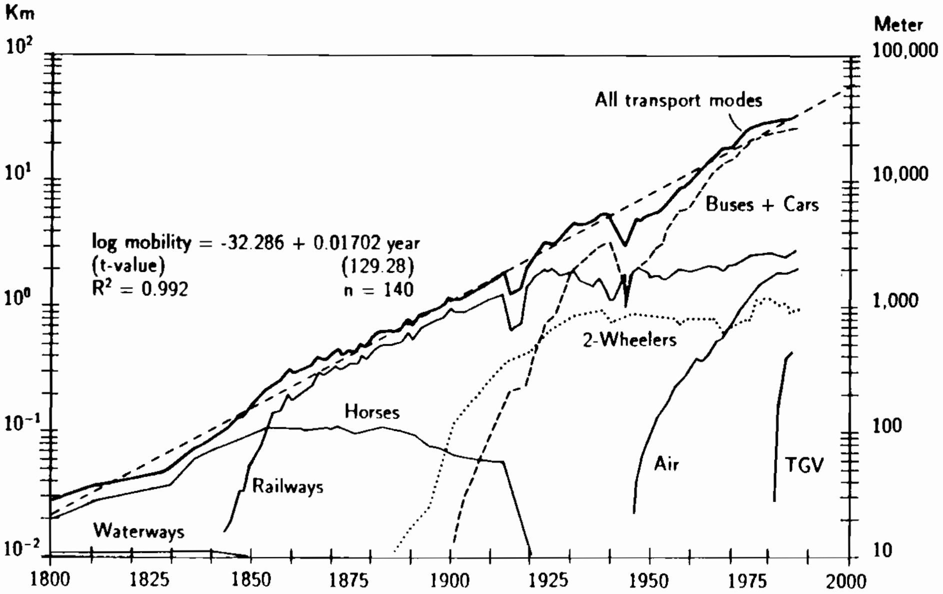


Figure 4.3.3. Range covered (average km travelled daily) per capita in France by mode and total since 1800.

total for all modes taken together since 1800. Each new transport mode increased the average range of an individual by about one order of magnitude. What is perhaps even more noteworthy is the absolute scale of the increase in travel range: from about 30 meters per capita per day (10.2 km per year) in 1800 to some 34 km per capita per day (12,400 km per year) in 1987, i.e., by a factor of over 1,000. Figure 4.3.3 shows a persistent exponential growth path in the individual daily travel range since 1800.

This trend toward higher mobility does not show any apparent indication to slowdown* or to "decouple" from its long-term growth trajectory. The situation in other countries is similar (Eberlein, 1988), although the distances covered vary in absolute terms due to differences in geography, spatial structures and the utilization of different transportation modes, as shown in Table 4.3.1. As a result of the increasing passenger-km transported by automobiles and aircraft (including travel abroad), at present the average distance covered per

---

* Note that the slowdown since 1970 has brought the mobility indicator just back to the long-term trend line.

Table 4.3.1. Average distance traveled daily (all modes[1)]), in km per day per capita in 1984-1986 for selected countries.

	Total	Including long-distance	Including by car and air
France	34.0	n.a.	28.1
FRG	28.4	12.8	23.6
Norway	30.1	n.a.	25.7
Sweden	31.8	n.a.	26.3
Switzerland	41.4	n.a.	34.7
USA	59.2	32.9	32.4[2)]
USSR	n.a.	10.4	2.0[2)]

[1)] Excluding walking and bicycling.
[2)] Only long-distance travel.

Source: Norway, Sweden: Schipper *et al.*, 1989; Switzerland: BA Stat., 1987, and IRF, 1986; all other: this study.

capita ranges between 30 to 60 km per day in OECD countries, half of which are for intercity (long-distance) travel.

Faster transport speeds therefore led to an expansion of the activity range of the population, geographical specialization of the economy, changes in urbanization patterns, etc. However, transport speeds and the resulting increases in the human activity range are certainly not the only contributing factors. Other factors, such as social acceptance and societal values, differences in the degrees of comfort of different transport modes, and last, but certainly not least, institutional and organizational settings (see e.g., the organizational history of railways) contributed to the historical development pattern of transport infrastructures. Still, we consider the (interrelated) variables *speed* and *range* to be key elements in the long-term evolution of the transport system, especially for long-distance travel, and premium market segments such as moving people and high-value goods.

High-value market niches

Into which segments of the transport market does a new technology and infrastructure diffuse first? An empirical analysis suggests that the (price insensitive) highest value market niches are conquered first, despite the fact that transport costs of a new system are initially much higher. This was the case with railways (conquering first passenger

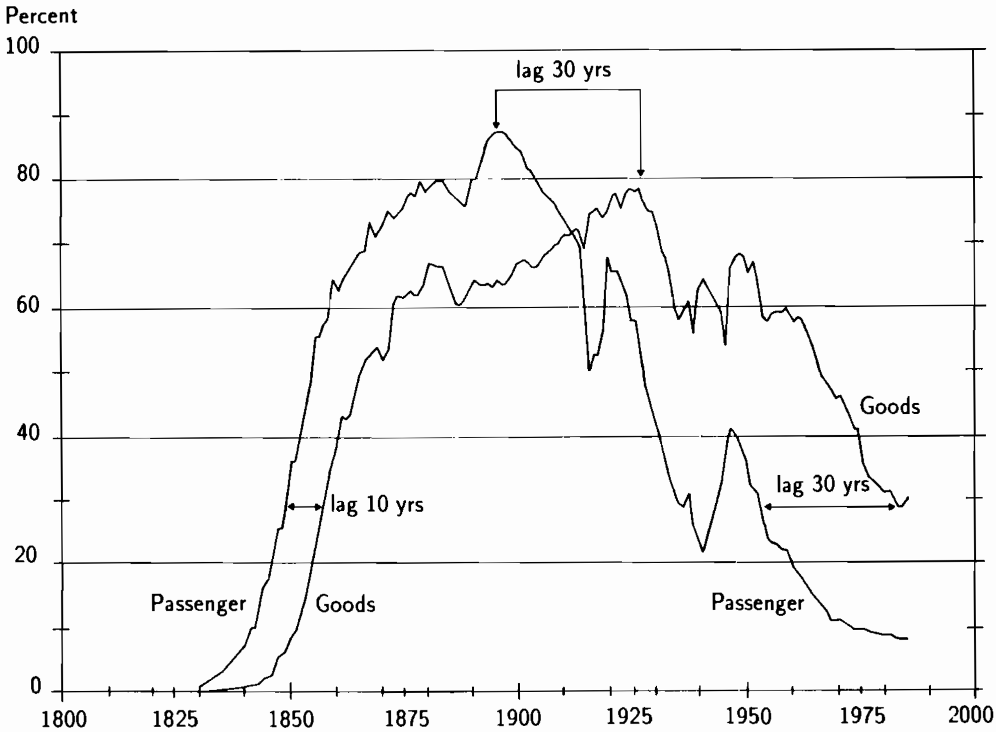


Figure 4.3.4. "First in, first out". Market share of railways during their life-cycle in high-value (passenger transport) and lower-value (goods transport) market niches, France 1830 to 1985.

and mail transport), the automobiles, and aircrafts. We would like to call this mechanism the "first in-first out" rule, in describing the importance of premium markets for the early development phase of new transport modes. These premium markets are represented first of all by long-distance passenger transport and secondly by high-value light goods, such as information (mail), luxury products, etc. Figure 4.3.4 illustrates this rule for market dynamics using as an example the life-cycle of a transport mode: the share of railways in passenger and goods transport in France.

The data for Figure 4.3.4 are derived from the long-term modal split for passenger and goods in transport in France, discussed in Section 4.2 above. Especially noticeable is the time lag in the market dynamics of the increasing share of railways between the premium and lower-value market segments. Railways first displaced other transport modes for moving people and, with a lag of around 10 years, followed a similar market penetration pattern in the goods transport sector. The maximum market share of railways was reached around 1900, with close to 90 percent of all passenger-km travelled, i.e., a *de facto* technological monoculture. A similar peak in market dominance in the goods transport sector was attained only some 30 years later with close to 80 percent of all ton-km transported by railways. In symmetry to the dynamics of gaining market shares, premium shares are lost first to new competitors. In France the decline trend in the market share of railways in passenger-km precedes a similar decline trend in ton-km by some 30 years. The situation in other countries and with other transport modes follows a similar dynamic pattern as formulated by the "first in-first out" rule. Recall here the much faster diffusion of automobiles in public and private passenger transport compared to goods transport in the UK and France at the beginning of this century (discussed in Section 3.3). Similarly, air transport first started to move information in the form of mail and later on passengers, whereas the share of air transport as a means of transporting goods, at least if measured in ton-km transported, is still practically negligible. We are reminded here of the dynamics of the intercity modal split in the USA (Section 4.2); cars reached their maximum market share in intercity passenger travel (around 90 percent) in the early 1960s, whereas the share of road transport in intercity freight traffic continues to rise, accounting presently for some 25 percent of the market (US DOC., 1987, and 1989).

Thus, the relatively price insensitive premium market niches provide the initial growth potential for a new transport mode. Other market segments, in particular goods of progressively lower value, follow at a later stage when significant cost reduction along the learning curve allows for competition (via tariff reductions) in lower value market niches. Only the lowest value goods (basic commodities, and bulk raw materials and agricultural products) remain as the last competitive domain for traditional (low speed) transport modes. This was, and continues to be, the principal market niche for inland river navigation in all countries, and is progressively also becoming the case for railways.

Under such an assumption, the development of new rapid rail based passenger transport systems, aimed at gaining shares in a premium market niche, appears to be strategically correct. Improved competitive advantage is, however, not always primarily a technological issue. A number of European experts (e.g., Sammer, 1988) claim that considerable institutional and organizational potential exists for improving the performance (door-to-door travel time) of railways in long-distance passenger traffic and that these organizational measures (e.g., new schemes for nationally coordinated train schedules along the lines of the Swiss *Taktfahrplan*) should be taken prior to massive investments into upgrading traditional rail infrastructure. Such strategies for improving the performance of railways are potentially

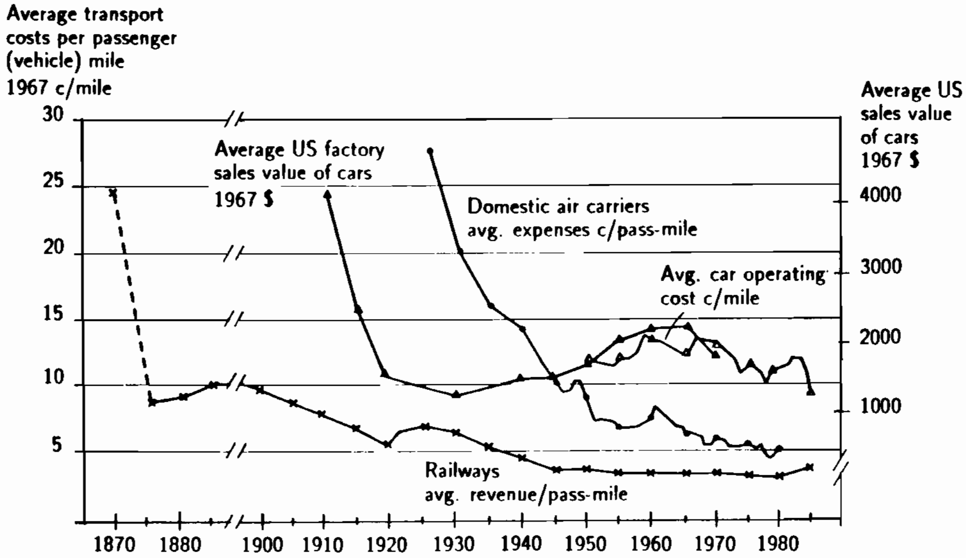


Figure 4.3.5. Learning curve in the average passenger transport mode costs in the USA, measuring average costs, consumer expenses, or revenues, in constant 1967 US cents per passenger (vehicle) mile and average factory sales value of new cars in constant 1967 US dollars. (Data source: Garrison, 1989, and US DOC., 1977, and 1987.)

successful only, if railways could become superior to private car or public air transport in door-to-door travel time. This becomes especially important, as in this premium market segment the lower transport costs of the railways did not noticeably affect their decline in long-distance passenger traffic.

Figure 4.3.5 illustrates for the USA the basic mechanism for proceeding from high-value market to lower value market segments: the learning curve. Initially high real term costs decline over time as a function of cumulative output. Despite initially significantly higher costs, each new mode of transport rapidly gains market shares, first in price insensitive market niches and later on (with declining transport costs) also in the total transport market. The tariff reductions proceed in high-value, price insensitive market segments (passenger transport) slower than in other market niches (e.g., goods transport) as was illustrated in Section 3.2 for the Canadian railway transport tariffs.

As no average total operating costs per vehicle-mile for private cars prior to 1950 were available for Figure 4.3.5, the time series were complemented by an important element of these costs, i.e., the fixed costs resulting from purchasing an automobile. The rapid decrease in the costs of an automobile, due to Henry Ford's Model T is particularly clear in Figure 4.3.5. It is also interesting to note that the in real terms the factory sale value of cars started to rise again after 1930, i.e., at the time when the first competitor, the horse, was eliminated from the market.

Initial costs per passenger-mile travelled were significantly higher for automobile (and air) transport than for other modes. Nevertheless, the car (and later the aircraft) started to diffuse rapidly and gained market shares. Even now rail prices (measured by the average revenues of railways per passenger-mile) are significantly lower than the cost of travelling by car. Assuming two passengers per car for long-distance travel, car operating costs are in the vicinity of five US cents (1967) per passenger-mile, i.e., still some 20 percent higher than railway transport costs. Thus, despite the fact that railways continuously enjoyed a transport cost advantage, their competitive position in long-distance passenger transport became progressively eroded, until railways virtually disappeared from this market segment.

Similar tendencies with respect to air transportation can also be outlined. Although initial air transport tariffs were higher than railway or car transport costs by as much as a factor of four, aircraft immediately started to gain market shares in long-distance passenger travel. As a result, the share of cars in intercity passenger travel started to saturate and decline as of the early 1960s. Air transport costs, however, approached similar real term passenger-mile costs as for cars, i.e., around five (1967) US cents per passenger-mile, only in the late 1970s. Despite higher costs the share of air travel has been increasing. With similar or lower costs than travelling by car, future market gains are even more likely. All this demonstrates that modal split decisions in premium market niches are highly price insensitive and point to travel time (speed) as being the most decisive preference criterion.

As transport tariffs become progressively lowered, further market niches are opened. We mentioned in Section 3.4 that the relative cost of exporting cars from Japan to the USA by aircraft compared to sea transport have been reduced already to a factor of two. We have also mentioned the large-scale export of semi-finished car bodies from Italy to the USA by aircraft. This implies that in the future air transport could become significant for the long-distance transport of lower value (per unit of weight) products, in addition to computers, flowers, fresh vegetables, etc., considered traditionally as small (measured by tonnage) special market segments of air transport. Further air transport cost reductions in real terms are by no means impossible. The trend of average air fares in the USA, presented in Figure 4.3.5 above, is a

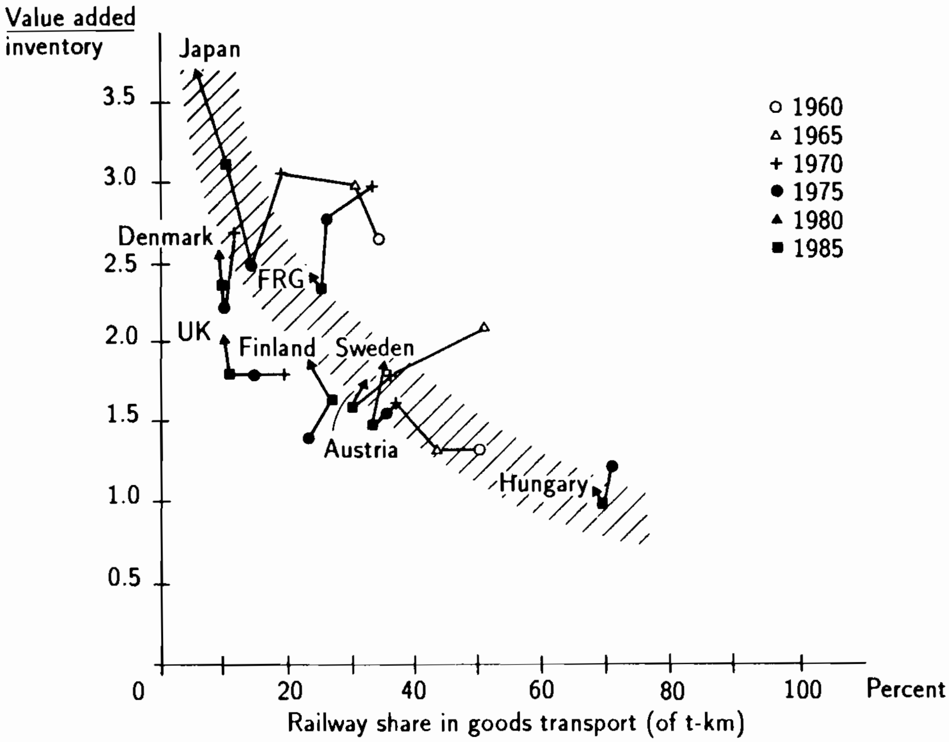


Figure 4.3.6. Ratio of value added over inventories in the manufacturing sector of selected countries versus railway share in goods transportation 1960–1985. (Data source: UN, 1977, and 1978, and IRF, 1970 to 1986.)

good indicator of the rationalization (cost cutting) potential of air carriers, especially when under strong competitive pressure. The rapid increase in fuel costs (constituting about one-third of the total operating costs of an aircraft) as a result of the two crude oil price hikes in the 1970s is barely discernible in the declining trend of average air passenger-mile fares. With further real term tariff reductions, which could become a reality under a progressively deregulated air transport market in Europe also, entirely new market niches could open up, possibly including regular long-distance commuting and large-scale goods transport.

The tendency for the latter to occur does not appear unrealistic, especially if the widespread application of the "just-in-time" principles in manufacturing and further efforts to reduce capital intensive inventories are carried out. The quest for inventory minimization and higher turnover in the manufacturing sector is illustrated in Figure 4.3.6 for a number of European countries and Japan. As an indicator of the performance of the manufacturing sector we take the ratio of the value added over inventories. A value of one, as is the case for

Hungary, implies that a whole years' worth of value added is tied up in inventories, either at the place of production, with the wholesale and retail dealers or in transit. On the other hand, a high-value of three, as for Japan, indicates that the total value added in manufacturing is turned over three times per year. In order to allow for warehousing and inventory minimization, and an increase in the turnover of working capital, speedy and reliable delivery is expedient. Thus the performance of the transport sector may explain some of the variances in the turnover rate in the manufacturing sector of different countries. We use the share of railways as an indicator of the transport system speed because railways are generally slower than road transport, and because high tonnage railway cars represent larger shipment sizes.

Although the relationship between both variables changes discontinuously over time, the turnover performance of the manufacturing sector and the performance of the transport systems appear to converge. In particular, this is reflected by the percentage of goods *not* transported by railways along the performance frontier shown in Figure 4.3.6. Countries like Hungary, where around 70 percent of all goods are still transported by rail, have the slowest (and most capital intensive) turnover rate. Japan, with the smallest share of railways in its goods transport sector performs best as regards logistics optimization (inventory and resulting capital minimization) in its manufacturing sector. Figure 4.3.6 records another indicator of the structural dilemma of railways. When production is geared toward higher-value goods, the industry will consequently aim at maximizing turnover and reducing the amount of time these goods sit around in railway cars waiting to be transported to their final destination. This poor performance of railways is not only the result of extremely low average transport speeds (very often in the vicinity of those of the preindustrial era, i.e., below 20 km/hr) but also of institutional and organizational settings which result in long standing times at various railway stations (sometimes in the order of several weeks, as for instance in Italy).

We conclude therefore that in addition to being the driving forces for future turnover improvements in the manufacturing sector, inventory minimization and just-in-time production regimes require a higher quality of service in terms of high transport speed, fast pick-up and delivery, and smaller batch sizes. All this should favor truck transportation or even air freight – or a combination of both – rather than rail. For instance already 50 percent of the air freight leaves Vienna airport but by truck (so-called "trucking").

Dematerialization and consequences for transport

We appear to be at the brink of a paradigm shift with respect to the long-term developments in goods transportation. Advanced economies are heading progressively toward increasing dematerialization (Herman *et al.*, 1989) with a higher software and information (and thus value) content in the products they manufacture. Figure 4.3.7 illustrates this for the USA. The material intensity of the US economy both in per capita as well as per constant unit of GNP generated, has passed its maximum and is on the decline. Similar tendencies can be shown for all advanced economies. This constitutes a major structural change and discontinuity in long-term economic development patterns. For the first time since the onset of the industrial revolution, per capita consumption levels and economic activity no longer appear to be tied to increasing resource inputs. This is certainly not the result of resource scarcity, but instead the consequence of a longer term structural change in the output mix of advanced economies. Increasingly higher value (and lighter) goods, embodying a higher information and software content, and service activities become the principal value generating activities of "post-industrial" societies.

This increasing "dematerialization" tendency can also be observed in the capital stock of advanced economies. Ever larger shares of capital investment are embodied in non-material information or software as illustrated in Figure 4.3.8. It is perhaps ironic, but nevertheless factual, as the history of the development of (especially air transport) technologies shows, that the development trends in the military sector tend to lead similar developments in civilian applications. As much as 80 percent of capital investment in the military sector in the USA in 1980 was in the form of (non-material) software (Ayres, 1989a). Software is projected to account for as much as 50 percent of capital investments into new manufacturing technologies, such as CIM (computer integrated manufacturing) and possibly more than 10 percent even in traditional manufacturing domains.

The emergence of the so-called "service economy" also points in the direction of increasing dematerialization of advanced economies. The service sector is already the largest consumer of "high-technology" in the form of computer hardware and software in the USA. New technologies in the service sector, whether automation in the New York Stock Exchange (Keith and Grody, 1988) or engineering such as bridge construction (Murillo, 1988) are considered to contribute to significant productivity improvements and further economic growth of the service sector (Guile and Quinn, 1988). Whether we consider employment or value (GNP) generated, the service sector

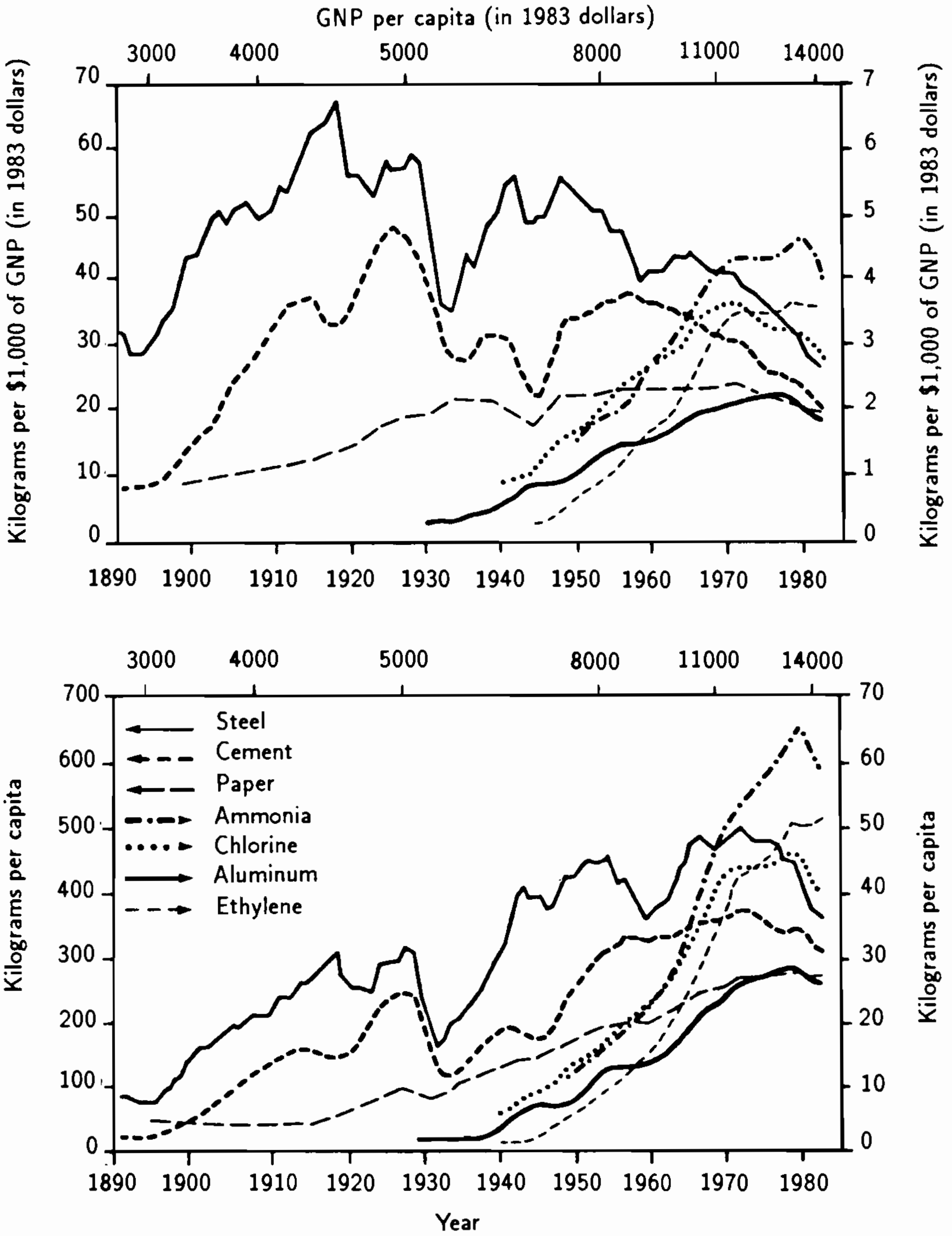


Figure 4.3.7. Trends in the material consumption per capita (bottom) and per constant (1983) US dollars of GNP (top) in the USA. (Source: Williams *et al.*, 1987.)

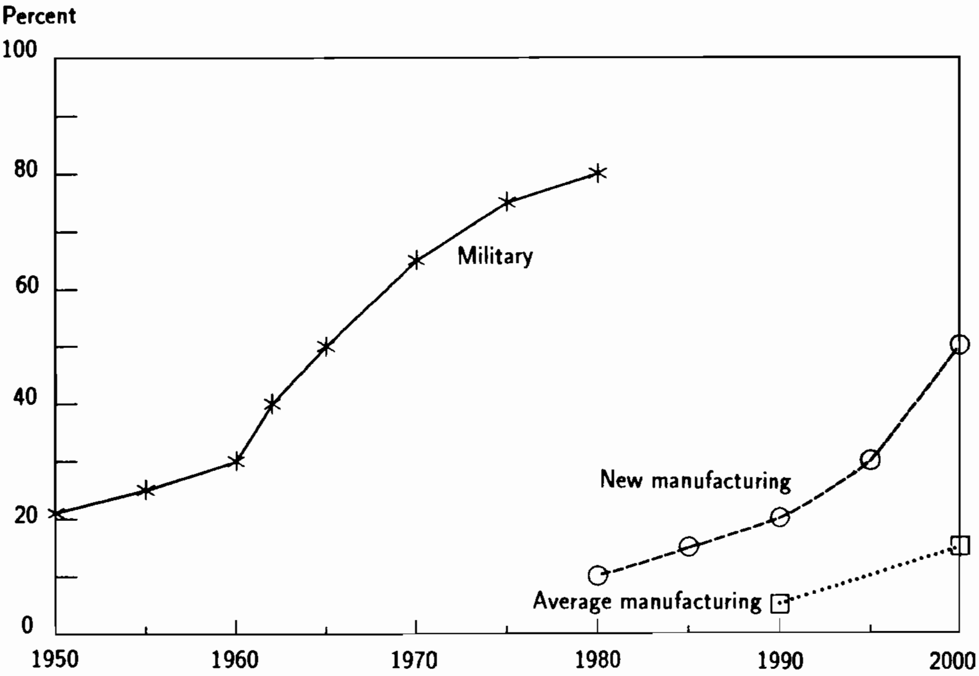


Figure 4.3.8. Software as a fraction of capital investment in military, new manufacturing and average manufacturing in the USA. (Source: Ayres, 1989a.)

will, without doubt, increase in importance in developed economies. This will have important repercussions on the transport sector in general. The service sector will impose new qualitative requirements on the transport sector, for moving people, information, and high-value goods instead of iron ore, scrap, and metallurgical products. Even if the precise nature of the output mix of advanced economies under such a major "paradigm shift" (to the "information and communication Kondratieff" in the sense of Freeman and Perez, 1988) is at present still uncertain, various functional characteristics of such a forthcoming new technological and socio-organizational "paradigm" have been conjectured. Recall here a few of the principles of our present industrial production organization: Fordist mass production, wide diffusion of consumer durables, high per capita consumption levels with resulting high energy and materials intensity, increasing international division of labor and importance of multinational enterprises. Freeman and Perez, 1988 (p. 53), describe the new functional characteristics of the forthcoming new production "paradigm" as follows:

> Diseconomies of scale and inflexibility of dedicated assembly-line and process plant partly overcome by flexible manufacturing systems, "networking" and "economies of scope". Limitations of energy and materials intensity [are] partly overcome by electronic control systems and components. Hierarchical departmentalization overcome by "systematization", "networking" and intergration of design, production and marketing. "Networks" of large and small firms based increasingly on computer networks and close co-operation in technology, quality control, training, investment planning and production planning ("just-in-time") etc. "Keretsu" and similar structures offering internal capital markets.

Here we only emphasize the principles of "economies of scope", decreasing energy and material intensity, as well as the integration of design, production, and marketing. All this implies "just-in-time" organization of design, production, and final delivery, with essentially zero inventories. The elasticity of demand for transport services in the future will therefore most likely be driven by increasing quality requirements as a result of the restructuring of advanced economies in the direction of higher value goods and services. In the following illustrative examples we show that different demand (modal split) elasticities of high-value goods and services in the transport sector can already be documented.

Figure 4.3.9 illustrates the different elasticities of various transport modes with respect to their respective (cumulative) share in the volume of (export) goods of different value transported in Sweden. Over 90 percent of the tonnage transported by railways and ships is worth less than 4,000 Swedish Kronas per tonne. Truck and air on the other hand, transport more than 60 percent of their tonnage in the value category above 4,000 Kronas per tonne. Thus, we can conclude that in the Swedish economy the largest share of high-value goods is transported by truck or by air. If the trend in the production of higher value goods continues, it can be expected that all this additional production will be shipped by truck or even by air, and that rail and inland or oversea navigation will be left with decreasing volumes of low-value goods. Thus, high-value goods appear to be particularly sensitive with respect to the transportation mode in which they are carried. The higher the value of goods, the (exponentially) lower is the share of railways and sea transport.

These findings for Sweden are further corroborated by an analysis of the elasticity of transport demand (share in the value transported) versus the average *value density* of the good transported, as illustrated in Figure 4.3.10 for goods imported into the FRG in 1986. In 1986 imports of manufactured goods into the FRG accounted for 424 billion

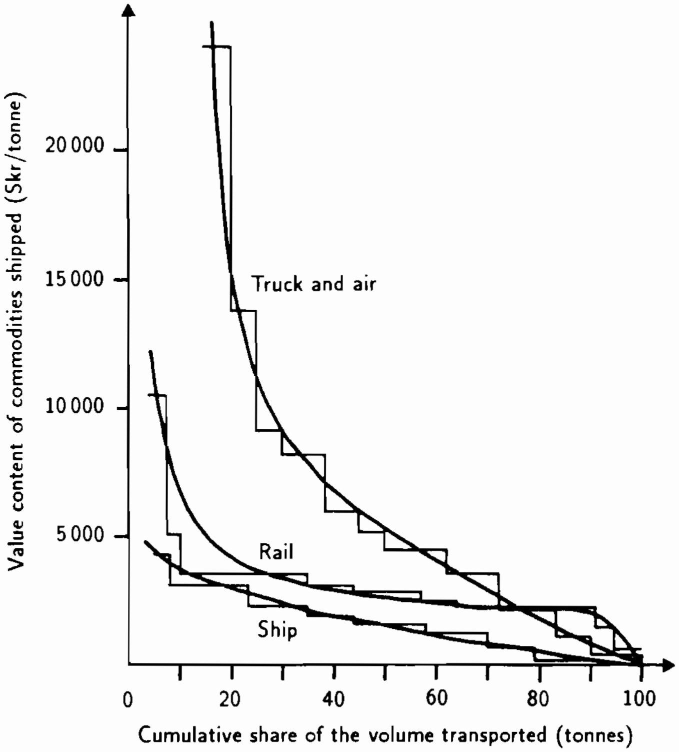


Figure 4.3.9. Value content of commodity shipments by three different modes of transport for Sweden in 1980. (Source: Snickars, 1987.)

DM for a total tonnage of 349 million tons (i.e., an average value of 1.2 DM/kg). Out of the import value, 50 percent arrived by truck, 24 percent by sea or waterway vessel, 10 percent by air, 7 percent by rail, and 6 percent by pipeline, (the remaining 3 percent being unaccounted for). It is interesting to note that while trucks transported 50 percent of the import value, they account for only 18 percent of the imported tonnage. For air transport this relation is even more extreme: 10 percent of the import value represent only 0.1 percent of import tonnage, resulting in an average value density of air freight of 166 DM/kg. Therefore, the importance of the various transportation modes changes significantly as a function of the *value* of the products shipped.

Figure 4.3.10 illustrates that basic materials such as coal, gravel, scrap, and raw materials in the value range below a few DM per kilogram are mostly transported by sea, canals, and rail (the dedicated oil pipelines are excluded in the analysis due to the limited possibilities for intermodal substitution). Inland navigation (canal) and sea transport account for around 25 percent each of the import value when dealing with the lowest value goods of one DM/kg. Railways account

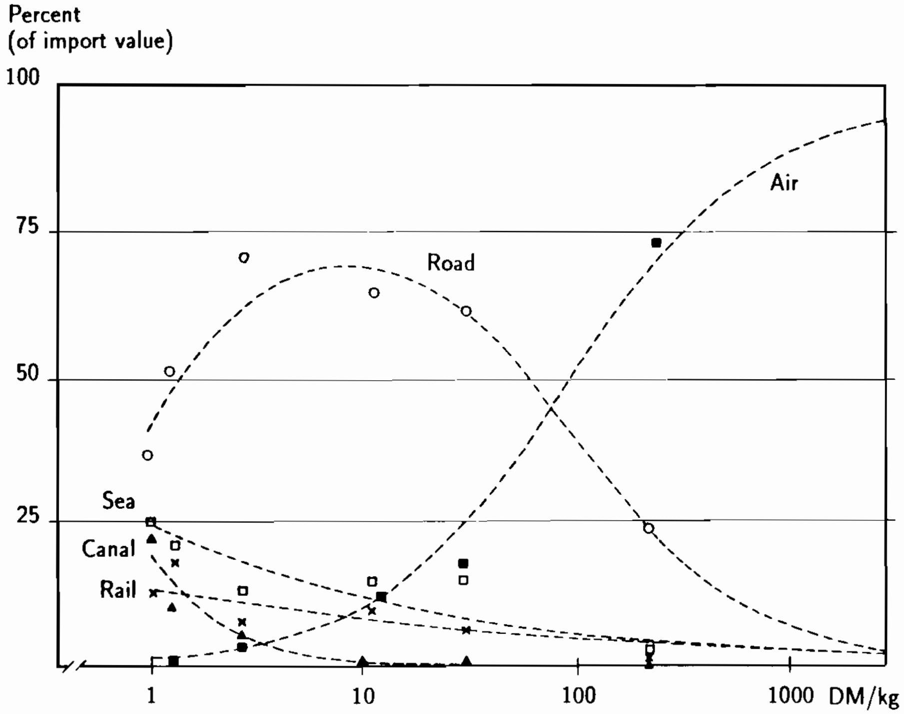


Figure 4.3.10. Share of different transport modes in the import value of the FRG in 1986 versus average value density of goods imported. (Data source: Stat. BA. Fachserie 7, 1986, and Stat. BA. (WiSta 7/1987), 1987.)

for more than 10 percent with road transport accounting for the remainder. Goods worth around 10 DM/kg are practically no longer transported by inland navigation. Thus canals appear the most (negatively) elastic with respect to the value of goods transported, a proof that inland water navigation is basically confined to transporting low-value basic commodities. Sea, and to a smaller extent also rail transport, are negatively elastic with respect to the value of the goods transported. In the category of 10 DM/kg, railways, sea, and air transport each account for about 10 percent of the import value, with the largest part (around 70 percent of import value) of these products being transported by trucks. Most manufactured goods, such as automobiles or machine tools fall into a value density range of up to 100 DM/kg. For this category, rail and sea transport nearly disappear as transport modes. The share of goods worth around 100 DM/kg that arrive by air and by truck is approximately equal, whereas railways and ships combined account for less than 10 percent of the import value in this relatively high-value goods category. Higher value densities, i.e., goods with values exceeding 100 DM/kg, such as electronics, computers or precision instruments, are usually shipped by air. Incidentally, the highest value manufactured goods (excluding precious metals, drugs or caviar) are aerospace products and aircraft themselves, all exclusively transported by air. The FRG, as illustrated in Figure 4.3.10, is not a special situation. Similar analyses for export or import relationships in Sweden, France, and Austria yield identical results. In fact, at the level of a cross-sectional analysis Figure 4.3.10 shows a similar dynamic for transport mode substitution as the long-term (longitudinal) infrastructure and modal split substitution patterns identified in Sections 4.1 and 4.2 above. This points to the self-similarity of the process of change, i.e., a fractal type pattern in the dynamics of change in the transport sector.

Under the growing significance of high-value goods production and just-in-time production regimes, the importance of air transport will grow substantially. Fast turnaround times as a result of high transport speeds, flexibility, and quality of service (around the clock pick-up and delivery) will be important performance criteria imposed on future transport systems. Specialized (private) fast delivery companies such as Federal Express in the USA consider themselves as the "largest warehouse in the country", pointing to the importance of immediate "just-in-time" delivery of goods between geographically widely spaced sites for design, production, and marketing of material and "non-material" goods (e.g., software, information). Plans by Federal Express for renting several Concorde airplanes for the transport and delivery of goods were based on rentability calculations in which the price charged per kg transported exceeded the price of a first class trans-atlantic passenger Concorde air ticket (per kg transported) by a factor of three. Thus, a significant market already exists in which the price for expedient goods transport significantly exceeds the costs we are apparently ready to incur for moving human beings faster over the Atlantic.

These may be of course considered as extreme examples, but they are early indicators of the possible developments emerging from further dematerialization and the generation of higher-value products by advanced economies. Such dematerialization tendencies would also appear desirable from the viewpoint of resource inputs and adverse environmental impacts of present material and energy intensive economies.

4.4. Long-Term Transport Intensities of an Economy

The long time series available on the transport output for France since 1800, discussed in Section 4.2 above, allow us to consider the long-term relationship between transport output and economic activity, i.e., ton- and passenger-km transported per unit of (constant) GDP. Figure 4.4.1 presents these three indicators: the constant GDP measured in 1913 French Francs (Toutain, 1987), as well as the total ton- and passenger-km transported. We observe that both ton-km as well as passenger-km have been growing at a faster rate than constant GDP. The growth rate of ton-km in relation to constant GDP has however been considerably smaller than the total passenger-km GDP growth. Since the 1970s, the rate of growth in ton-km no longer proceeds faster or even parallel to GDP growth, but is considerably slower pointing at a decoupling of goods transport from GDP growth.

Figure 4.4.2 analyzes the resulting transport intensity per constant unit of GDP. Note that the turbulent years during World War I and II, with their resulting unreliable GDP and transport output estimates, have been omitted in Figure 4.4.2. Ever since the 1930s and especially after World War II, the passenger transport intensity increases more sharply than that for goods. Both variables show that from a long-term perspective, the *transport intensity* of the French economy (and we would conjecture similar tendencies in other countries also) has been steadily rising. The goods transport intensity increased by a factor of 4 between 1800 and 1985, i.e., from 0.197 to 0.79 ton-km per 1913 Franc GDP, corresponding to an average growth rate of 0.75 percent/year. Passenger transport intensity has, on the other hand, increased by a factor of over 53 (!) during the same time period (i.e., from 0.04 to 2.2 passenger-km per 1913 Franc GDP). This corresponds to an average growth rate of 2.2 percent/year. Considering that moving people represents the highest value, premium market segment of the transport sector, the much higher historical growth rates of passenger transport intensity should not come as a surprise. The increase in transport intensity is in contrast to energy intensity, which, as a historical analysis for the USA since 1800 has shown, has been decreasing* at an average annual rate of about 1 percent per year.

One should not be biased by the high figures for transport output per unit of constant GDP, as the GDP data are expressed in constant

---

* Source: Nakicenovic, 1984. Energy intensity in the USA has been declining ever since 1800. The energy input of the US economy (including also non-commercial energy forms such as fuel wood and feed for animal power) has declined from 13 KWyr per 1958 US dollar GNP in 1800 to around 2.2 KWyr per 1958 US dollar GNP at present, i.e., by nearly a factor of 6.

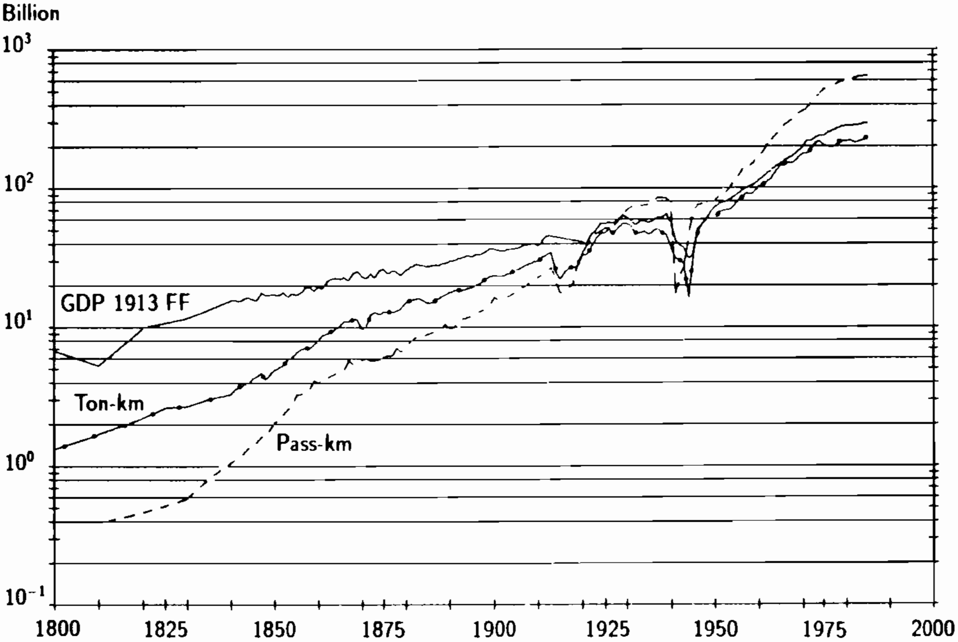


Figure 4.4.1. Transport output and constant GDP in France since 1800, in billion (10^9) ton- and passenger-km transported and billion 1913 French Francs.

1913 money terms. In terms of current French Francs (FF) the transport intensity in 1985 corresponded to around 135 passenger-meters travelled and to around 47 kg-km transported per 1985 FF GDP in 1985. Assuming 90 kg per passenger, passenger transport corresponds to around 12 kg-km per FF (1985) GDP, and total traffic thus to around 60 kg-km per 1985 FF GDP. Transport and communication accounted for around 6 percent of GDP in 1985. The increasing transport intensity of the French economy observed in Figure 4.4.2 does not imply that an even higher share of the GDP is spent* in the transport sector, at least not in a similar fashion as the increase of transport intensity per constant unit of GDP would suggest. Recall here, the drastic real term transport cost decreases in the case of the USA (Figure 4.3.5). Although no detailed statistics for France are available, one can assume a similar decreasing tendency (the real term railway tariff reductions in France were mentioned in Section 3.2 above).

---

* Toutain, 1987, estimates that transport accounted for some 1.7 percent of the *produit physique* in 1830 and for around 3 percent in the 1930s, compared to 6 percent presently (including, however, also communication).

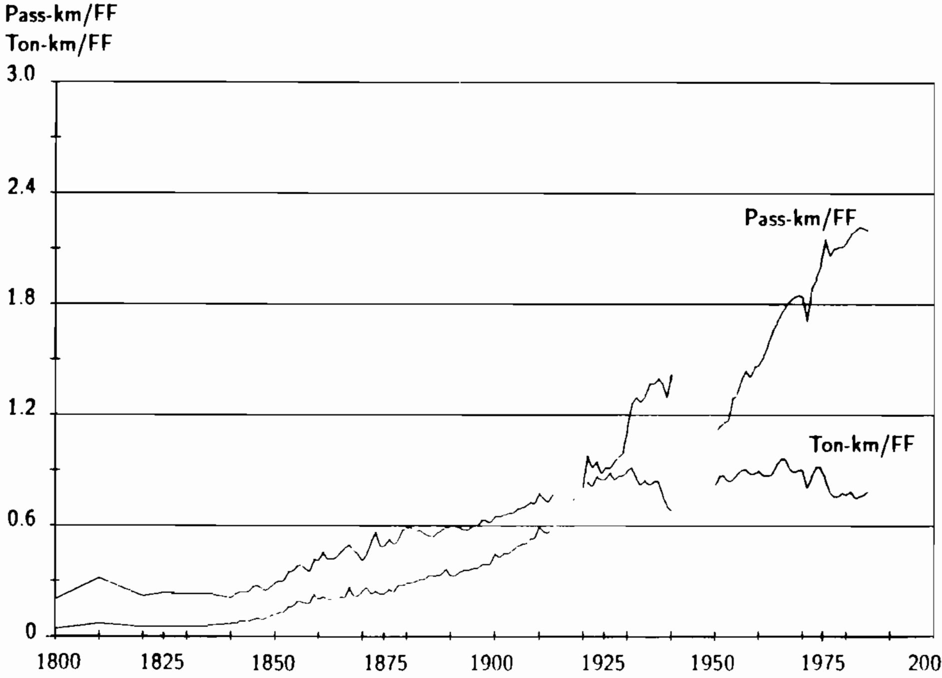


Figure 4.4.2. Transport intensity of the French economy since 1800, in ton-km and passenger-km per 1913 French Franc GDP.

The drastic decrease in real term transport costs certainly stimulated further increases in the transport intensity of the French economy, as outlined in the discussion of the UMOT transport model (Chapter 1), which puts forward empirical evidence that on average a constant fraction of 15 percent of disposable family income is spent on transportation.

From Figure 4.4.2 it becomes clear that the transport intensity of goods passed its maximum in the 1960s and appears to have been declining ever since. We have noted that this might constitute an important discontinuity in the long-term development of goods transport, resulting from the structural shift to increasing dematerialization and higher shares of (low weight) high-value goods and services in the output mix of developed countries in their transition to "post-industrial" economies. Thus, it appears likely that the ton-km transported per unit of GDP will continue to gradually decline in the medium-term future.

However, similar tendencies are not apparent in passenger transport intensity. The latter has been increasing since the onset of the 19th century, first due to the introduction and growth of the railways,

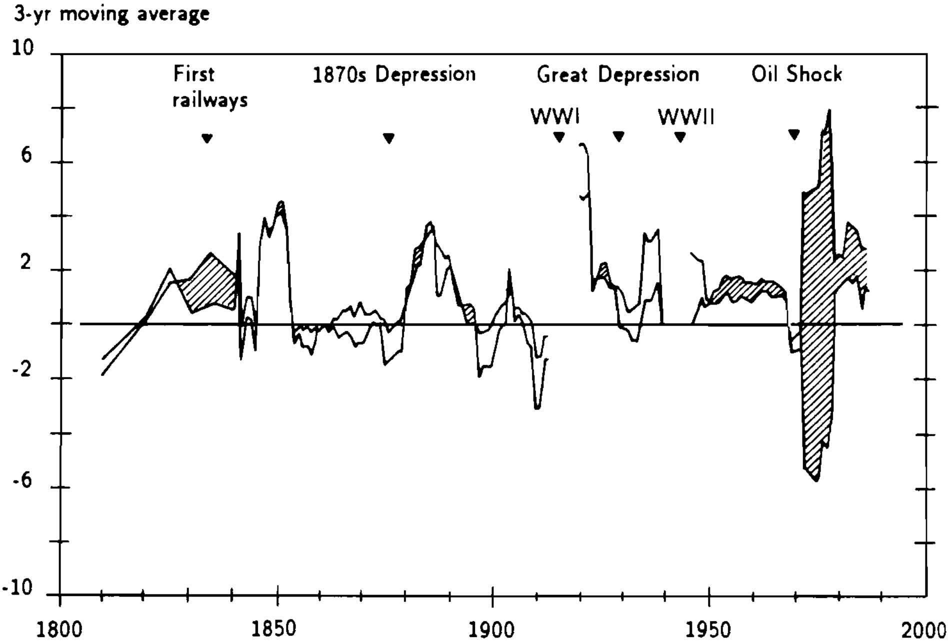


Figure 4.4.3. Total passenger- and ton-km to constant GDP growth elasticity in France, yearly elasticities smoothed by a 3-year moving average.

later the automobile and recently aircrafts. As in the discussion of the long-term passenger modal split in Section 4.2 above, we note a major discontinuity as a result of World War II. The passenger- and ton-km growth in relation to GDP growth discussed above, may also be analyzed in terms of the commonly used indicator of elasticity, i.e., the relationship of transport output growth to economic growth as shown in Figure 4.4.3. The war-time period elasticities are omitted due to the inherent uncertainties of GDP and transport output estimates during World War I and II.

A first observation from Figure 4.4.3 is that over most of the time period under consideration the elasticity is significantly above 1. Second, it shows typical fluctuations as a result of medium-term business cycles (the very strong annual fluctuations in elasticity are smoothed in Figure 4.4.3 by a 3-year moving average). It is noteworthy that up to the 1970s, both passenger and goods transport elasticity with respect to GDP growth, never bifurcated, i.e., that while being different in various historical phases, both showed a synchronous behavior and moved in the same positive or negative direction. The prevalence of passenger over goods transport elasticities is indicated by the shading in Figure 4.4.3. Unshaded areas indicate time periods where the good transport was growing faster than passenger transport per unit GDP growth. Historically we may conclude, that as a rule goods transport growth reacted more sensitively to business cycle variations. Since 1950 a divergence in the passenger and goods transport elasticities to real term GDP growth can be identified. Passenger transport growth was consistently above the goods transport growth per unit of growth of GDP. The higher passenger elasticity since 1950 in Figure 4.4.3 implies that passenger-km have grown significantly more per unit of real term GDP added than ton-km.

Since the early 1970s, however, we observe a complete bifurcation. Goods transport growth appears to progressively decouple from economic growth as shown in the high negative elasticities over an extended period of time as a response to the two oil price shocks. In contrast, passenger transport elasticities have been highly positive. Whereas this may be a specific case for France (a similar analysis for the FRG showed no such strong bifurcation), it is nevertheless an indicator that in future the transport demand for passengers and high-value goods may portray a different behavior to that of low-value products.

At present, it is not clear whether the historic trend observed in France is also representative of other countries, especially in the long-term. Still, it appears likely that with increasingly affluent societies, passenger transport demand will continue to grow in the future. In the goods transport sector, increasing affluence means an increase in the value and information (software) content, and a decrease in the material and energy input per unit value added of the goods produced and consumed. This implies that a further decoupling of ton-km transport volume from economic growth might be expected in the longer term. This decoupling does not mean of course that the total ton-km transported will decrease, but rather that its growth rate will fall behind that of GDP growth, since much of the economic growth will be based on the service sector with its "immaterial" (information and value intensive) type of products.

The growth trend of passenger-km transported per capita or per unit of constant GDP observed above, encourages us to speculate about the outlook for the future. The long-term growth trend of passenger transport intensity of the French economy was disrupted only once, as a result of World War II. We have observed that the French society and economy behaved *as if* it was a "hibernation" period. Long-term social and economic growth trends became temporarily reversed during the war, but reassumed the pre-war growth

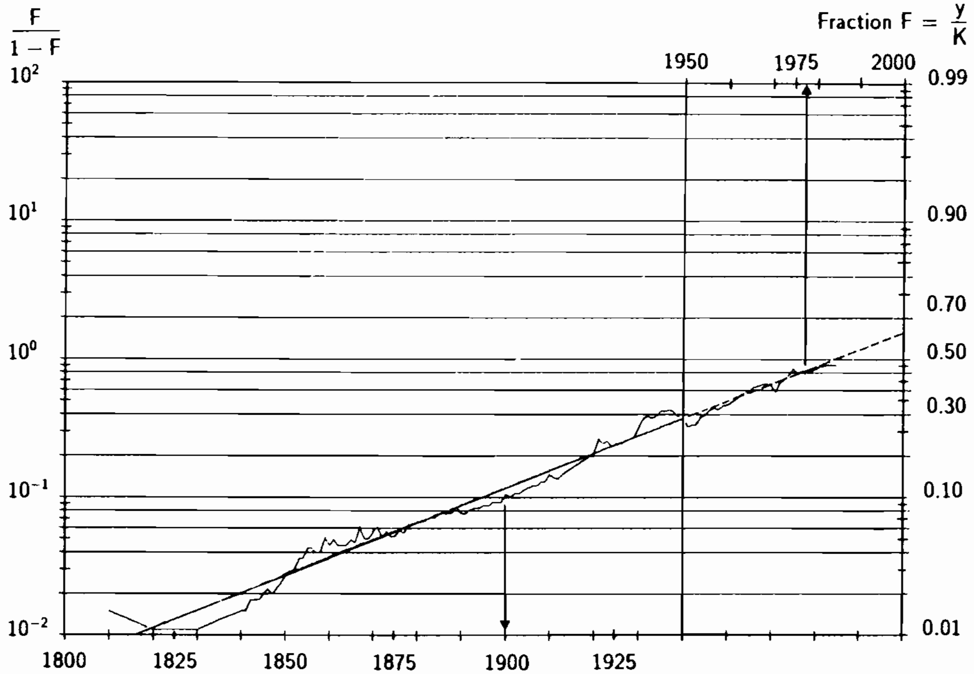


Figure 4.4.4. *Gedankenexperiment* in the analysis of the long-term growth trend in the passenger transport intensity of the French economy (in passenger-km transported per 1913 Franc GDP). The experiment assumes, that the time period between 1940 and 1950 (World War II and its aftermaths) never happened. Note in particular the split in the time axis between the bottom and top scale of the figure.

trend (rate) about 10 years later, once the pre-war diffusion or intensity level was reached again.

For such a speculative outlook for the future we propose the following *Gedankenexperiment.* What would be the long-term trend in the growth of passenger transport (intensity) assuming that World War II never occurred? The *Gedankenexperiment* suggests the removal of the period of "hibernation" between 1940 and 1950, by assuming that these years in fact never existed, i.e., we assume for our analysis a kind of "time reform" (thus the year 1951 becomes the new "synthetic" year 1941 and so on). Figure 4.4.4 reports the results of this *Gedankenexperiment.* Note in particular the discontinuity in the time axis. After 1940 (bottom of Figure 4.4.4) comes 1951 (i.e., 1940 and 1950 are identical on the split time scale) and from there on the data refer to the top part of the x-axis. The long-term time trend was estimated using a logistic function, primarily in order to avoid suggesting that the growth of passenger transport could proceed without any ultimate limit. The growth trend up to the present has been basically exponential (as reflected in the estimate of the growth function prior to its inflection point).

Despite the fact that we deal here with a hypothetical case, the continuation of the pre-war growth trend emerging from Figure 4.4.4 is striking. It is as if the passenger transport intensity, after a period of "hibernation" continues smoothly the dynamic trajectory it started 100 years before the major discontinuity of World War II. Based on this *Gedankenexperiment* it appears that we might be only half-way through the long-term growth trend of passenger transport intensity. Because the analysis is hypothetical, we refrain from presenting this result as a possible forecast.* Instead Figure 4.4.4 is intended as a simple observation of a stable long-term trend in the premium market segment of the transport sector: moving people. From this, we may conclude that at least the dynamics of the process (the *growth rate*) was apparently unaffected by a major societal and economic discontinuity. Under the continuation of such a long-term growth trend, passenger transport could well grow significantly** more in the future. Such a long-term growth scenario would, of course, still have to be corroborated by similar analyses for other countries. Unfortunately the available data allow only the consideration of a (possibly uncharacteristic) recent short time segment of the long-term growth trend in passenger transport. Our conjecture about possible future growth is, however, corroborated by an analysis of the long-term relationship between transport and communication, discussed in the following Section. From this analysis we derive that passenger transport and communication evolve as complementary, rather than substitutable economic goods. The possible emergence of new information and communication technologies may, under such a conjecture, provide for further growth in passenger transport. Is man a "territorial animal" as claimed by Marchetti, 1987, aiming to "control" ever larger "territories", travelling ever larger distances, as was apparently the case during the last 200 years?

---

* Therefore no model parameters and statistical measures of fit are presented. From a statistical point of view, the uncertainty in the parameter estimation is in fact so high (as the growth process has not reached the inflection point yet), that one cannot conclude whether the process is approaching any long-term saturation level at all or is simply following an exponential growth path.

** Assuming a growth rate of the real term GDP of two percent per year, passenger-km could increase by up to a factor of 6 over the next 70 years, based on the above *Gedankenexperiment*.

4.5. Substitutability Versus Complementarity of Transport and Communication

In Section 4.4, we have conjectured that the historical trend in the increase of passenger transportation over the last 200 years is likely to continue in the medium-term future. We have referred to the fact that a number of authors consider that the future direction of economic growth will be progressively shaped by a new "paradigm" in the organization of production, summarized under the heading of the "information and communication Kondratieff" (Freeman and Perez, 1988). Such a hypothesis is corroborated by the increasing tendencies toward dematerialization and higher information and software content in products as well as the further growth of the service sector in advanced economies.

If one of the predominant features of a next step in the evolutionary development of our economies is the widespread application of new information and communication technologies, a most natural question arises as to how this would affect the postulated continuation in the growth of passenger traffic. It has been advanced that the widespread diffusion of new information and communication technologies could result in a significant reduction of personal travel, i.e., that new communication technologies will eventually substitute for travelling.

In the discussion of the long-term evolution of passenger transport it was pointed out that the introduction of new systems are of particular importance in the long-term growth of total passenger transport volume and the structural changes in the relative shares of different transport modes. The same can be said for communication technologies. Both the speed of communication as well as the type, form, and information content of messages exchanged were drastically altered by new communication technologies introduced in the past; first the telegraph then the telephone, and possibly in the future by new communication systems, intergrating video-phones, telefaxes, and information and data exchange via interconnected computer networks.

Thus, as a first step in our analysis, we examine the relationship between different modes of communication in a historical context. The conclusion from such an analysis suggests that in the area of communication, a sequence of replacements progressively changes the morphology of the communication system and contributes to overall system output (number of messages exchanged) growth. The introduction of the telephone was possibly the most fundamental and important innovation in the recent history of communication. Figure 4.5.1 suggests that in the case of France, the market impact of the telephone (measuring its share by the number of messages exchanged)

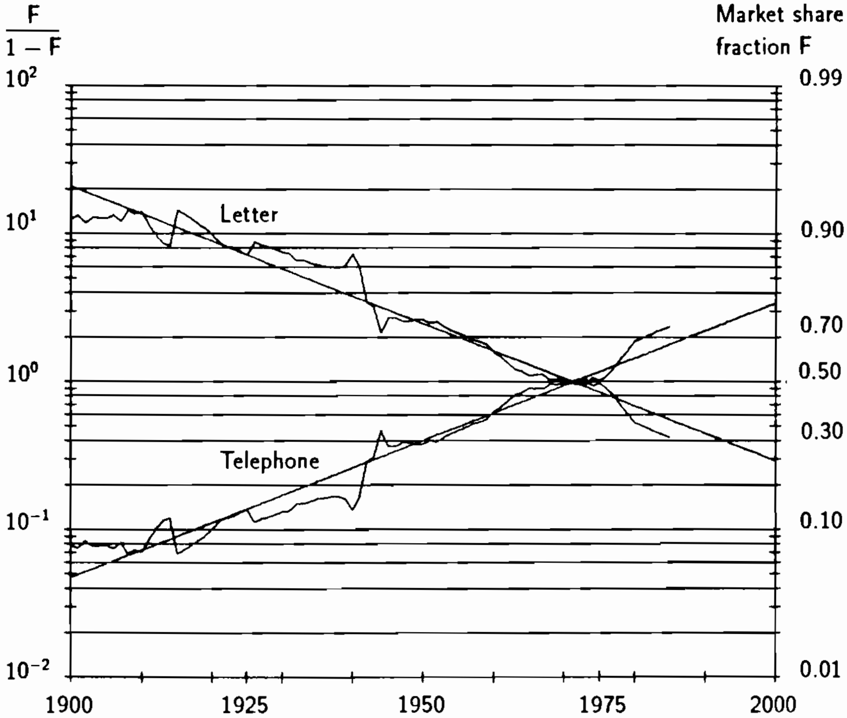


Figure 4.5.1. Substitution of letters by telephone calls in the total number of messages communicated, France 1900 to 1985.

while being significant, was certainly not of a rapid short-term nature. In analyzing the share of letters and telephone calls (both domestic and external communications) in the total number of messages communicated in France, we see a long-term structural shift (substitution), which appears to proceed with a Δt of 53 years and t_0 in 1971. All other forms of messages, telegrams, faxes, telexes, etc. account together for only between one and two percent of all letters and telephone calls, and have been consequently excluded from Figure 4.5.1. The impact of the telephone on communication behavior as represented by our crude* measure is thus very long-term, since even in the early 1970s half of the messages communicated in France were in the form of letters. Analyses for other countries show an identical pattern (Batten, 1989).

---

* Our quantitative measure of considering the share of letters and telephone calls in the total number of messages transmitted ignores of course important qualitative differences, which may exist in the form and content of messages transmitted (e.g., a love letter compared to a business call). A similar problem occurs in the area of passenger transport performance (e.g., a daily commuting trip in a crowded train as opposed to a weekend excursion by bicycle to the countryside). However, such qualitative characteristics cannot be captured within the quantitative analysis framework followed here.

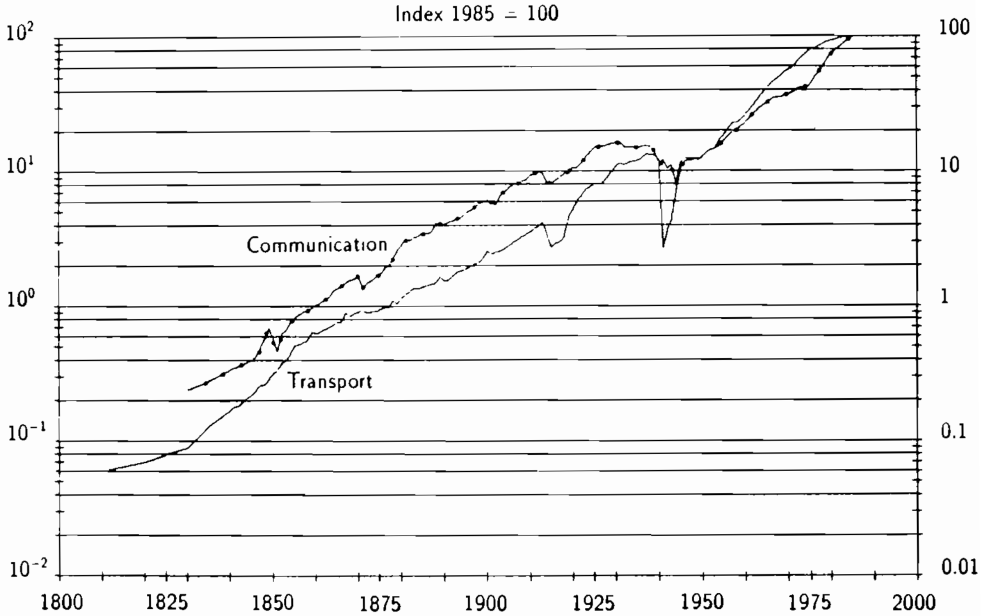


Figure 4.5.2. Growth of passenger transport and communication in France, measured in passenger-km and total number of messages transmitted in form of an index 1985=100.

In order to examine the long-term relationship between passenger transport and communication, we compare the total output of the two systems in France over a long period in history. This allows an analysis of whether the introduction of different new transport and communication systems has changed the relationship between them. Figure 4.5.2 reports the growth of total passenger-km transported and total output of the communication sector, i.e., total number of messages sent, by letter, telegram, telex, telefax, and telephone (data source: Mitchell, 1980, and Ann. Stat., 1961 and 1987) since the beginning of the 19th century. The two measures, not directly comparable, are normalized by using an index of 100 for their respective output in 1985. Both transport and communication have increased by over four orders of magnitude since the beginning of the 19th century. The impacts of both world wars, in particular World War II are clearly visible in Figure 4.5.2 and have been discussed already in some detail in the preceding Chapters. It is visible from the piecewise linear (exponential) growth trends in Figure 4.5.2, that from a historical perspective, transport and communication have evolved in unison.

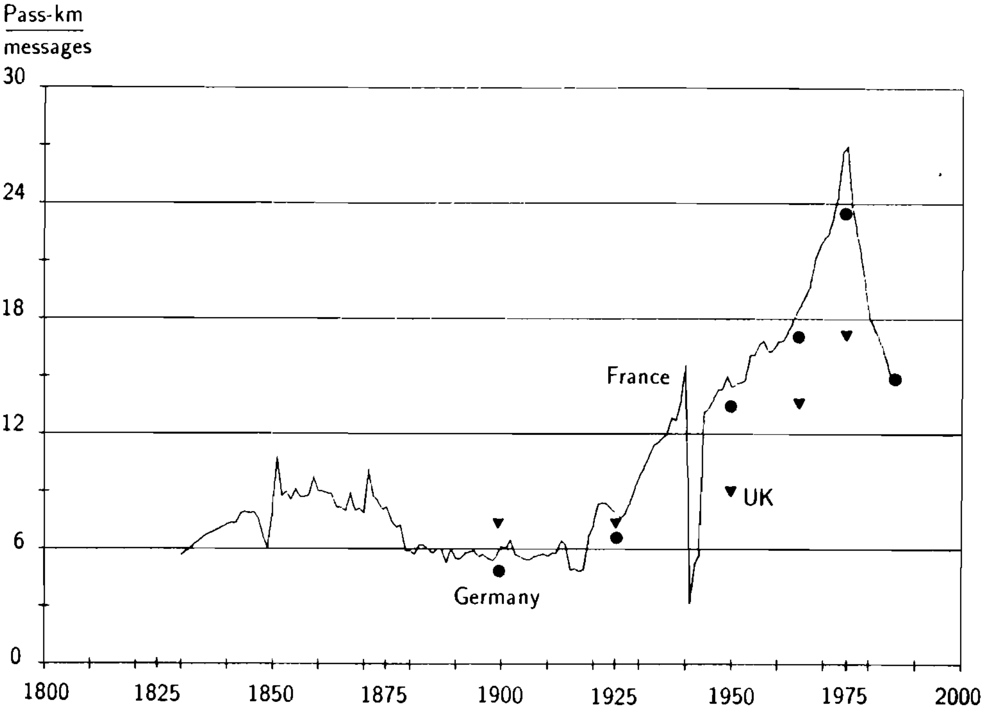


Figure 4.5.3. Ratio of passenger travel to messages sent in France, Germany (FRG after 1945), and the UK. (Data source: Mitchell, 1980, and IRF, 1960 to 1986.)

Any long-term substitution tendency between transport and communication would only appear if the exponential growth paths would diverge significantly over longer time periods (as reflected in different slopes). Figure 4.5.2 indicates however, that both transport and communication grew at similar long-term rates, i.e., they evolved as complementary goods rather than substitutable ones.

However, transport and communication have not always grown at exactly the same rate. Passenger transport has grown faster than communication over a period of several decades particularly due to the introduction of railways and later automobiles. In the long-term, however, the two seem to reestablish parity. This long-term parity rate appears to be in the vicinity of around six passenger-km travelled per message transmitted. As shown in Figure 4.5.3, this is not only true for France but also for Germany and the UK. Thus, one might conclude that whatever growth potential due to new communication technologies realized, it may just be sufficient to bring the ratio of face-to-face to information channel communication back to long-term parity. Thus, while the growth of transport and communication did not evolve completely in unison, they do enhance each other and show many parallel developments. New communication technologies may consequently be seen as "catch-up" instruments in the long-term complementary and synergistic relationship between transport and communication, a hypothesis at least consistent with 200 years of history.

Although a number of (e.g., business) trips may be substituted by new communication technologies, it is equally important to consider that new communication technologies may in turn also *induce* additional travel, as argued for instance by Cerwenka, 1986.* Particularly under a given travel time budget constraint, it appears possible that any trip savings due to new communication technologies, whilst reducing the passenger-km travelled for a specific purpose, such as business, may induce trips in other domains such as leisure or holiday travel.

The various, differentiated effects of new communication technologies on different travel purposes (work, education, business, shopping, leisure, and holidays) are comprehensively discussed by Cerwenka, 1986. For our purposes here it suffices to report his conclusions: due to the diffusion of new telecommunication technologies there will be no reduction in passenger travel, instead considerable growth is likely to occur. Thus, *travel and communications are better seen as interrelated elements in a social context which they help to create* (Albertson, 1977).

---

* Cerwenka, 1986, reports estimates that new communication technologies may reduce business travel by an amount corresponding to around 0.5 percent of the total passenger-km travelled in the FRG. Cerwenka and Rommerskirchen, 1983, report that presently already as much as 48 percent of all passenger-km travelled in the FRG are for leisure and holiday purposes. This percentage is expected to increase to over 50 percent in the near future.

CHAPTER 5

Infrastructures and Long Waves*

> *The history of the productive apparatus is a history of revolutions. So is the history of transportation from the mailcoach to the airplane. The opening up of new markets, foreign or domestic, and the organizational development from the craft shop and factory to such concerns as U.S. Steel illustrate the same process of industrial mutation – if I may use this biological term – that incessantly revolutionizes the economic structure from within, incessantly destroying the old one, incessantly creating a new one. This process of Creative Destruction is the essential fact about capitalism.*
>
> Joseph A. Schumpeter,
> *Capitalism, Socialism and Democracy*, 1942

Economic growth is not a continuous process. Periods of high growth rates in output and of stable evolution in the form and direction of economic growth can be identified in economic history, like the *Belle Epoque* of the 19th century or the post World War II boom ending at the beginning of the 1970s. Economic theory argues that the economy is expanding along an equilibrium growth path. Such periods do not, however, last indefinitely; periods of growth and expansion in economic activities are punctuated with phases of fundamental changes in the structure of the economy, the technological base, and many social institutions and relations. Some of the most important changes in socio-institutional frameworks and in the economic structure since the onset of the industrial revolution are related to the pervasive adoption of new systems, including for instance the development of infrastructures. Today, the emergence of new systems such as

---

* Parts of this Chapter are drawn from research carried out at IIASA together with Nebojsa Nakicenovic within the framework of diffusion and long-wave research. Some of these results have been published elsewhere, see for example: Grübler, 1987; Nakicenovic, 1987; and Grübler and Nakicenovic, 1990.

Table 5.1. Clusters of pervasive technologies.

1770-1830	1820-1880	1870-1940	1930-1990	1980-2040
Water Power, Sails, Canals, Turnpikes, Iron Castings, Textiles	Coal, Iron, Steam Power, Mechanical Equipment	Railways, Steam Ships, Heavy Industry, Steel, Dyestuff, Telegraph	Electric Power, Oil, Cars, Radio, TV, Durables, Petrochemicals	Gas, Nuclear, Aircraft, Telecomm., Information, Photo-Electron.
Mechanical Equipment, Coal, Stationary Steam Power	Steel, City Gas, Indigo, Telegraph, Railways	Electricity, Cars, Trucks, Radio, Phone, Oil, Roads, Petrochemicals	Nuclear, Computers, Gas, Telecommunication, Aircraft	Biotechnology, Artificial Intelligence, Space Industry & Transport
Manufacture	Ind. Prod.	Standardization	Ford-Taylorism	Quality Control

Source: Adapted from Pry, 1988.

information and communication technologies is often mentioned in this context because there are reasons to believe that they may become *pervasive* throughout the economy, i.e., that they will not diffuse in just a few selected sectors.

The growth of pervasive systems entails a large set of technological, infrastructural, but also institutional and organizational innovations. For example, the diffusion of motor vehicles was contingent on the development of numerous other technologies and infrastructures such as paved roads, the internal combustion engine, oil refining, motor fuels and oil pipelines, new sheet metals and high quality steels, electrical equipment, and a whole host of other new technologies and products. Table 5.1 illustrates this clustering and cross-enhancing effect of a number of interrelated technological systems, which have been driving forces for particular economic expansion and growth periods since the onset of the industrial revolution.

The list of technology clusters as presented in Table 5.1 is illustrative rather than exhaustive and also the timing of the various development periods is only approximate. Still it represents a general overview of the main technology clusters, which were responsible (frequently studied under the leading sector hypothesis by economic historians, e.g., as for railways) for much of the extent and the general direction of economic growth during particular historical periods. The top of Table 5.1 presents the principal technology clusters dominating a particular epoch of development. The lower part indicates technologies which are introduced and emerge during the same phase, in order to become pervasive in the next phase of development. For example, the dominant cluster of the 1930s to the present time includes: growth in electric power, oil and petroleum, roads and cars, petrochemicals, radio and TV, instruments and controls, etc. During the same period we had innovations and the emergence of new industries which may become important in the future: nuclear energy, computers, natural gas, telecommunication and information systems.

On the bottom of Table 5.1 we have attempted to summarize the general "trajectory" (Nelson and Winter, 1982) or principle driving force and direction of economic growth during the various epochs: from manufacturing to industrial production to the emergence of standardization, then to economies of scale based on Fordist production and scientific management principles (Taylorism), in the period from the 1930's to the present, and finally economies of scope, and quality control and management most likely in the next phase of development.

The transition from one phase of development to another is triggered primarily by the fact that the technological cluster, responsible for much of the economic growth, approaches a number of limits: saturating markets, diminishing returns from further incremental improvements and productivity gains, increasing awareness of social and environmental disbenefits associated with a continuation, and further intensification of the traditional mode of development. The "crises" in predominant "techno-economic paradigms" (Freeman and Perez, 1988) are generally manifested in increased turbulence and volatility in markets, prices, and even in social relations. At the same time, it is exactly during these transitional adjustment phases that clusters of new technologies emerge and the characteristics and direction of the next phase of development are being shaped by a social process. These changes in the techno-economic paradigms, illustrated by the emergence and diffusion of interrelated clusters of pervasive technologies, enables us (as shown below) to integrate the history of transportation into the more broader long-term changes in economic structure and social fabric.

Whereas there is wide agreement about the discontinuous nature of long-term economic growth and the importance of technological change, there is still controversy (see e.g., Rosenberg and Frischtak, 1983) surrounding the issue of whether the historically discontinuous development pattern involves long-term regularities, what are its causes, and whether there are associated recurrent fluctuations of relevant quantitative indicators such as prices, profit rate, employment, innovative activity, trade, investment, etc. Research has often concentrated on the extent to which the fluctuations in price levels are cyclical, whether there is synchronization, whether the variations are only limited to monetary indicators or can also be extended to physical measures of output, materials or employment. Below we attempt to describe the phenomenon of *Wechsellagen* in economic development as a process of structural change that goes beyond the question as to what extent prices and other monetary indicators might portray inflationary tendencies over a period of a few decades followed by periods of disinvestment. We document that periods of growth and expansion in economic activities, and in particular in infrastructure development, are interlaced with phases of fundamental changes in the technological and infrastructural base, the structure of the economy, and also many social institutions and relations.

To this end, we summarize below the empirical findings and the resulting long-term patterns in the development of transport infrastructures identified in Chapters 3 and 4. We have observed, that both *technology clusters* like the transport technologies based on the steam engine (railways, steamships, etc.), and *country clusters* (international diffusion bandwagons), i.e., countries participating in a major expansion impulse, as characterized by their similar diffusion pattern, tend to saturate during relatively short time periods. This short time interval, during which most diffusion processes become completed and eventually enter a phase of decline, is the more noticeable in view of the long time constants involved in the growth of large-scale systems. We have argued, that this "synchronous" saturation is the result of a "catch-up" effect in the diffusion pattern between countries, but also within a given technology cluster (e.g., the acceleration in the introduction of new technologies in the US car industry, shown in Section 3.3). Such "clusters" in the completion of a large number of interrelated technological diffusion and substitution processes in various countries have been referred to as the *season of saturations*.

Thus, long-term infrastructure development is seen to proceed through a *dichotomy* of expansion periods, characterized by the widespread diffusion of new technological systems as represented for example by the canal age, the steam (coal) and railway age, the automobile (oil) and road age, etc. Boom periods are followed by a clustered saturation of the predominant techno-economic features responsible for the previous upswing, including (at present the case with automobiles) the widespread realization of negative externalities resulting from the growth of the dominant technological regime. The *season of saturations* would, from such a perspective, indicate that some kind of economic, socio-technical, and environmental feasibility (or social acceptance) limits are being reached. "Gales of creative destruction" (Schumpeter, 1935, and 1939) prevail in this period of turbulence and volatility, but provide in turn possibilities for adjustments to new requirements and for the emergence of new socio-technical solutions. The latter, whilst (latently) already in existence, have been effectively barred from widespread diffusion during the previous growth phase (Brooks, 1988), due to the commitments to traditional infrastructures and capital vintages. Thus, a particular (interrelated) cluster of solutions determining the form and direction of economic growth is "locked-in" (Arthur, 1983) which hinders the emergence of new systems. Technical, organizational and institutional innovations and reforms (many of them accumulated over the past decades) may be turned into practical applications only once the traditional socio-technical regime enters crisis. The latter paves the way for overcoming existing feasibility limits (the "technological stalemate" of Mensch, 1975) and for transgressing the turbulent period of discontinuity in order to reassume a stable evolutionary path thereafter.

In essence, the above is already a verbal model* of long-term discontinuities in economic development as reflected in the evolution of the transport sector observed in the empirical part of our analysis. However, such discontinuities cannot be discussed in isolation for the transport sector alone. An analysis of the transport sector has to be embedded into the overall evolutionary pattern of the economy and of technological change in general. This is necessary in order to show that the empirical observations and the mechanisms responsible for them, are consistent with an important line of research on long-term patterns of economic growth: Long Wave (*Kondratieff wave*) theory.

Following Van Gelderen, 1913, Nikolai Kondratieff, 1926, and 1928, and especially Joseph Schumpeter, 1939, long-wave theory has received an increasing amount of attention in recent years. Empirical and theoretical work, building on the basis laid by these pioneers, has confirmed their conclusions about the discontinuous nature of long-term economic development. Studies of long-term fluctuations in economic development have grown to such considerable volumes, that an overview of the origins, lines of research followed, and empirical evidence assembled arguing *pro* or *contra* would require more space than is available here. Therefore, we refer only to some works that provide a good overview, in particular Bruckmann, 1983, Delbeke, 1983, van Duijn, 1983, Freeman, 1983, Goldstein, 1988, Kleinknecht, 1987a, and Nakicenovic, 1984.

Whereas empirical evidence indicates beyond doubt that long-term economic development has indeed proceeded through a succession of major discontinuities, the debate continues (see e.g., the papers contained in Vasko, 1987), as regards the cyclicity and recurrence of

---

* For a formal mathematical model along these lines see for instance Mosekilde and Rasmussen, 1986.

economic fluctuations, their nature (whether long waves are a phenomenon confined to capitalistic economies or extend even to periods prior to the industrial revolution), and especially with regard to their driving forces. To be concise, the debate centers on the prevalence of various causal factors of long-term economic fluctuations, in particular innovation and entrepreneurship (e.g., Mensch, 1975; Marchetti, 1980; Kleinknecht, 1986, and 1987a), capital and investment (e.g., along the Marxist tradition represented for instance by Mandel, 1980, or the systems dynamics studies of Forrester, 1983); employment (in connection with discontinuities in innovative behavior and different sectorial growth, as argued by Freeman *et al.*, 1982) or the availability of raw materials and foodstuffs (e.g., Rostow, 1975, and 1980). Other factors (for an overview see in particular, Goldstein, 1988) considered include migration flows (Berry, 1988) and macropsychological factors related to socio-technical change (De Greene, 1988).

Although many of the above mentioned factors are interrelated and much of the debate centers on the question as to which factor outweights the others in influence and importance, three main areas of agreement about long-term economic fluctuations emerge:

(1) Economic growth is not a smooth continuous process, but is characterized by major discontinuities and a (regular or irregular) succession of periods of extended growth with periods of retarded growth, recession or even depression.
(2) Empirical evidence suggests, that particularly in price series, significant discontinuities are revealed for most industrialized countries, in some countries even for periods prior to the onset of the industrial revolution i.e., before 1750.
(3) The importance of (technological, social, and institutional) innovations and discontinuities in innovative and entrepreneurial activities is acknowledged by most researchers.

The line of research into discontinuities in the *appearance* of innovations was influenced very strongly by the work of Mensch, 1975. Although there exists convincing evidence that the "bunching" in the appearance of innovations during periods of recession and depression is much weaker than originally maintained by Mensch, a certain clustering of innovations can still be (at a statistically significant level) identified. However, the "bunching of innovations", is not going to explain the whole sequence of discontinuities in economic development from the slowdown in growth to recession, recovery, and growth. Innovations, i.e., new technologies at their moment of introduction proper (a typical innovation in Mensch's sample is the Stephenson locomotive), will have only a marginal impact on the economy as a whole. An innovation "bunch" would only provide for economic growth at a later stage, when being translated (via *diffusion*) into the creation of whole new industries and product lines. Long-wave research has almost entirely concentrated on the *appearance* aspect of innovations (formulation by Mensch, 1975, vivid critique by Freeman, Clark and Soete, 1982, and first conclusion by Kleinknecht, 1987a and 1987b), whereas the *diffusion* aspect, i.e., how do innovations contribute to economic *growth* and – when saturating – to economic *recession*, has received limited attention. Despite the fact that most researchers agree about the theoretical rationale that economic slowdown and recession is a saturation phenomenon, extensive empirical evidence has yet to be assembled. Consequently, we try to fill this gap by analyzing empirically the diffusion history of a large sample of innovations in the case of the USA, covering the whole "life cycle" of innovations, i.e., their introduction, diffusion, and saturation. We show that discontinuities in the long-term evolution occur, even without a rigid clustering in either the appearance, growth or saturation dates of innovations.

In this context, innovations should not be defined only within a narrow technological context. Innovations consist of a host of interrelated technological, institutional, and organizational new ways to perform traditional or entirely new tasks, to produce traditional or entirely new products and services, etc. The importance of such a set of interrelated, interdependent technological, institutional, and organizational innovations, driving major economic expansion periods was convincingly argued by Dosi, 1983; Perez, 1983; and Freeman and Perez, 1988. Such an interrelated set of innovations has been referred to as technological "trajectory", "paradigm" or "socio-technical paradigm", by various authors. For our purposes here, it suffices to stress the importance of interdependencies between technological innovations proper, and the associated new forms of production organization, new institutions, management methods, etc., required to fully realize the inherent growth potential of a new technological stream.

Consequently, we integrate the evolutionary pattern of infrastructure development into discontinuities in the long-term development of the economy as a whole, in order to suggest that transport technologies and infrastructures constitute important "indicator" innovations for various historical socio-technological paradigms, the growth of which (due to their demand and investment stimulating effects) was responsible for economic upswing periods. The saturation of the predominant paradigm in turn results in a slowdown of economic growth and a phase of (economic) restructuring, characterized by periods of economic and social turbulence and volatility (as reflected in price data). It is during these periods of turbulence and structural transition, that the technological, economic, as well as the institutional and/or organizational characteristics of a new paradigm are shaped, which through its progressive diffusion, provides the basis for a new upswing.

We emphasize again, that we do not argue for a strict and rigid clustering in the growth and saturation periods of a large number of interrelated innovations. It is sufficient that a larger number of important innovations overlap to some extent in their main growth and saturation phases, whereas some technologies and market segments "tunnel through" the period of saturation and turbulence (the "Kondratieff barrier") in order become part of the main growth sectors in the next upswing period. This was the case with the continuing expansion of railways after 1870, automobiles after 1930 and is, in our opinion, currently the case with air transport.

Consistent with one area of agreement by most long-wave researches, the discontinuities in the long-term development of an economy (in our case the USA) is illustrated by two indicators: prices, represented by energy and crude oil prices* and a new measure of discontinuities of *innovation diffusion.*

One of the widely used indicators for identifying turning points in long-term *Wechsellagen* of economic growth is the fluctuation in price levels. Figure 5.1 shows real term average energy prices as well as the real term price of crude oil in the USA since 1800. The energy price trends shown in Figure 5.1 are quite representative of the general price levels in the US economy, as reflected for instance in wholesale prices (Nakicenovic, 1984). Despite many turbulent events and acute crises, and the complete transformation of the USA from an agrarian to an industrial society over the last two centuries, the long-term energy price trends have been surprisingly stable with the exception of four pronounced flares: prior to 1820, around 1860-1870, the 1920s, and in the 1970s and early 1980s.

Figure 5.1 demonstrates that the price flares since the OPEC oil embargo have been deflated back down to historical levels. Thus, in spite of the recurring brief episodes of extreme price volatility, crude oil (and energy in general) has shown rather remarkable price stability in real terms over its more than a century-long history as one of the major world commodities. With the exception of three prominent price flares around the 1870s, prior to the Great Depression and after 1973, the price of crude oil fluctuated within the relatively narrow

---

* Data source: US DOC, 1975, and 1980 to 1987. The average real term energy price data are based on price series by energy carrier, weighted by their respective market shares in the primary energy balance of the USA.

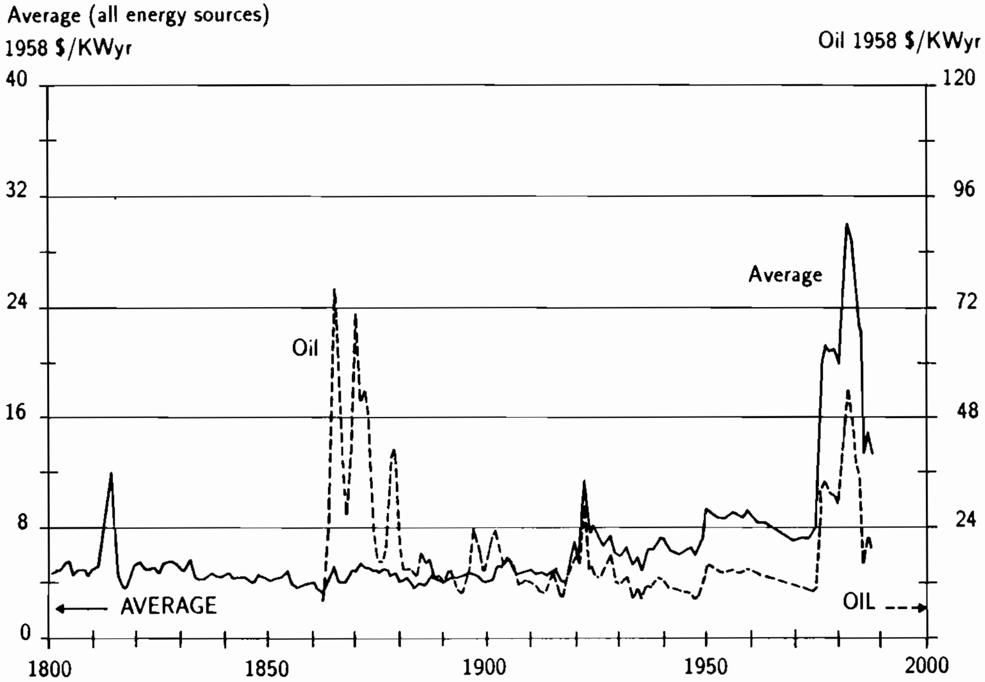


Figure 5.1. Real-term oil and composite price of all energy sources in the USA since 1800, in 1958 US dollars per KWyr.

range of about US $13/bbl (in 1987 US dollars, or about US $18/KWyr* in constant 1958 US dollars as shown in the figure). It has been shown elsewhere (Nakicenovic, 1988, and Grübler, 1987), that periods of price volatility mark important turning points in the structural evolution of the US economy in terms of the saturation of predominant technological systems as well as in the introduction of new ones.

Energy prices accurately indicate** rather than trigger, an important structural discontinuity in the long-term evolution of a system. These discontinuities in the energy sector are mirrored by the changes in energy technologies and their infrastructural base, represented by the emergence of new primary energy forms and saturation in the market share of the dominant energy carrier (Nakicenovic, 1984, and Grübler and Nakicenovic, 1987): emergence of coal around 1815; saturation of feed as the primary energy carrier and

---

* Kilowatt-year per year (31.5 10^9 Joule). One KWyr equals about 5 barrels (bbls) of crude oil.

** At this point a possible analogue to our body temperature can be made. Rather than being the cause, a rise in body temperature is an *indicator* of important physiological changes (e.g., an infection).

emergence of oil around 1870; saturation of coal and emergence of natural gas (not produced as a byproduct from oil fields) around 1920; and the present saturation of oil and the emergence of nuclear energy. The energy price series reconstructed here, also allow a comparison of expenditure on energy in an economy compared to GNP. Apart from the periods of price volatility, primary energy inputs to the US economy account typically for two to three percent of GNP. The fact that this ratio has not increased over the last 200 years (except for limited periods of market volatility) suggests that, at least in the energy field, resource scarcity phenomena can *not* be observed in the long-term.

Against the background of the rather static indicator of real term energy prices (despite the four pronounced flares around the stationary trend), we document below the dynamic changes in the replacement of old by new pervasive systems. As a measure of long-term discontinuities we will consider the *diffusion of innovations.* Here we are interested in analyzing the long-term tendencies in the *rate of technical, economic, and social change.* For this purpose, we have taken a sample of 117 innovation diffusion cases for the USA, complemented by a second sample comprising 265 diffusion histories. Details on the methodology as well as on data sources used to estimate the parameters of the diffusion models are contained in Grübler, 1988, and will not be repeated in detail here. The basic idea is to analyze the average rate* of technical, economic, and social change by averaging the *first derivatives* of the diffusion and/or substitution functions of a large number of processes of change, which are derived from long empirical time series. To clarify, this is the diffusion equivalent to the annual GNP growth rate.

Innovations, consistent with the concept of socio-technological paradigms presented above, are not only taken from the technological field alone. The empirical cases considered include the areas of energy, transport (i.e., some of the results presented in this work), manufacturing, agriculture, consumer durables, communication, military, and finally economic and social diffusion and structural change processes (e.g., diffusion of literacy, reduction of infant mortality, structural changes in employment, etc.). The characteristics of our sample may be summarized as follows: Diffusion rates differ widely between different innovations. Measured by their Δt, i.e., the time required to go from 10 to 90 percent of ultimate (growth or market share) potential, the diffusion rates range from 2 to 300 years, with the mean in our data sample being 57 years.** Diffusion processes,

---

* This is the sum of the growth rates for all diffusion processes in a given year, divided by the number of observations.

** 41 years in the second sample of 265 innovation cases. The second sample includes in addition to the IIASA sample many short-term technological change processes, the parameters of which were not estimat-

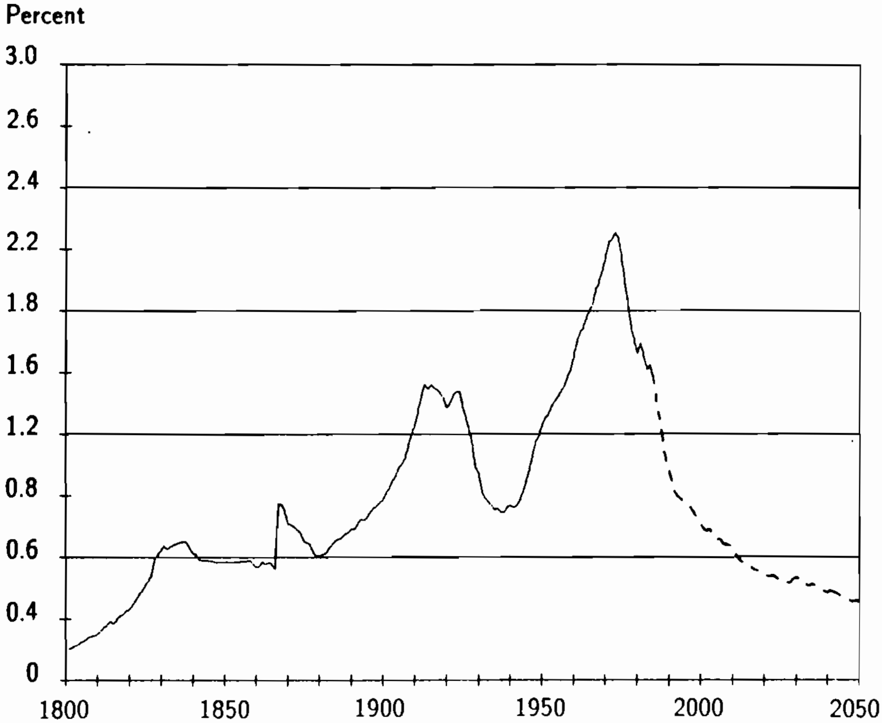


Figure 5.2. Average rate of technical, economic, and social change in the USA, based on the diffusion of a sample of 117 innovations, in percent/year.

measured by their time constant Δt exhibit a rank size distribution. No general acceleration tendency of diffusion rates over time has been observed in our data samples. The only acceleration tendency noticeable is the one towards the end to the main growth period, i.e., similar to the acceleration tendency towards the "Kondratieff barrier" observed in this study. The mean diffusion rate of 57 years is a possible indication that the famous 55-year Kondratieff cycle may indeed be closely linked (we conjecture caused) by the average rate a society and an economy absorbs (discontinuously) new innovations. Figure 5.2 summarizes the results of this analysis by showing the aggregated *rate of change*, i.e., the average of the first derivatives of all diffusion processes occurring at a particular time period.

The analysis reveals clear peaks and troffs in the average diffusion rate, i.e., the average pace of change in the US economy and society. Long-term development is therefore far from being a continuous process. The longer-term trend indicates an increase in the

---

ed ourselves but obtained from a large number of diffusion studies (referred to in Chapter 2).

aggregate level of diffusion. This could, however, be the artifact of the sample. There are simply more diffusion processes documented for the last hundred than the previous hundred years. Although we are taking averages, the higher number of overlapping processes in one interval could result in a higher aggregate diffusion rate. In weighting the large number of different diffusion processes (see discussion below) this increasing tendency in the average rate of change is indeed shown to stem from the data sample, confirming however, the conclusion with respect to the discontinuous nature of socio-technical change.

Each peak characterizes the beginning of the saturation of the corresponding cluster or family of diffusion processes. The reason why the curve tapers off to the right of the graph (after the present time) is that it could be assumed that some diffusion processes with very long Δts will continue in the future. Presumably many innovations have emerged during the last decades that may turn out to be successful, and if they were included perhaps we could expect a trend reversal in the rate of change curve sometime after the mid-1990s. Such an inherent lag is due to the slow diffusion rates at the beginning of the introduction phase of innovations, i.e., the flat initial part of the S-shaped diffusion curve. The turning points in the evolution of this new empirical measure of the long-term rate of change are consistent with the turning points of economic development identified by a number of long-wave researchers (e.g., van Duijn, 1983; Goldstein, 1988; Marchetti, 1983; Nakicenovic, 1984; and Vasko, 1987).

Our working hypothesis is that the diffusion processes within successive clusters or families of interrelated systems enhance each other and promote the pervasiveness of each of the successive development trajectories. A classification or a taxonomy of each individual innovation, required to make a whole technology cluster viable, would probably reveal a hierarchical system with one successful diffusion having a positive feedback and catalytic effect on the development of many others within the whole cluster. We assume that it is the interdependence of individual diffusion processes that in time focuses each of the clusters in order to approach saturation within a relatively short period of time. Thus, the increasing rates of change in Figure 5.2 indicate the diffusion of a whole technology cluster, which we can associate with the development of individual transport infrastructures: canals up to 1860, railways up to the 1920s, and roads and automobiles up to the 1970s. The declining average rate of change that follows the peaks, indicates a gradual transition of many diffusion processes to a phase of saturation and onsetting decline, i.e., a *season of saturations.* During this period, the growth potential (rates) of the dominant paradigm becomes progressively exhausted, whereas the newly emerging paradigm, (although already partly in existence) is still in its very early phase of diffusion, with corresponding small growth rates.

Of course our analysis has up to now just considered a large number of innovation diffusion cases, irrespective of their macro-economic effects. If one does not want to resort to a subjective, and thus necessarily arbitrary, ranking of a large number of diffusion processes, a theoretical framework for weighting various diffusion processes ought to be developed. We suggest that a classification or a taxonomy of diffusion processes along a hierarchical system might be a way to assess the different macro-economic effects of different diffusion processes.

To this end, Table 5.2 shows a hierarchy of diffusion processes taking into account their temporal dimension, i.e., a hierarchy along the required diffusion time between the introduction and the saturation of diffusion processes in the transport sector, taking into consideration the different diffusion rates, Δt between the USA and the USSR (Tzarist Russia prior to 1917). The reason for considering two countries with entirely different market structures relates to the validity of such a hierarchical classification of diffusion processes. We consider, that if the processes can be classified in a hierarchical structure, despite the economic differences between a market and a centrally planned economy, that any derived weighting measure will be robust across different systems (countries) and by analogy also between different diffusion processes (i.e., over time). In addition to the diffusion time constant Δt, Table 5.2 also shows the inflection point t_0 of the diffusion process, i.e., the time of the maximum growth (replacement) rate at the 50 percent diffusion (market share) level.

A hierarchy of diffusion processes all along from the macro down to the micro level emerges when considering the time constants involved in the growth (diffusion) of transport systems and of technological change (substitution) in transportation. Another interesting finding, which emerges from Table 5.2 is that while the diffusion processes are shifted in time (time lag between t_0 in Table 5.2), the diffusion time constants appear to be of a similar order of magnitude between the two countries. Of course decisive differences also remain as for instance the second phase in railway construction in the USSR compared to a decrease in the size of the network in the USA (discussed in Chapter 3 above).

The similar duration in the diffusion of technologies and infrastructures between the two countries, enables us to develop a rough hierarchical classification of diffusion processes, as reflected by their diffusion constants Δt. Between each of the hierarchical levels of temporal diffusion processes, the diffusion constant Δt varies approximately by a factor of two. At the very macro level, when considering

Table 5.2. Hierarchy of diffusion processes in the transport sector of the USA and the USSR, measured by their temporal diffusion parameters.

	USA		USSR	
	t_0	Δt	t_0	Δt
Total Length of Transport Infrastructures	1950	80	1980	80
Growth of Railways				
Network 1830-1930	1858	54	1890	37
Railways 1930-1987	decline	decline	1949	44
Treated Ties (USA)	1923	26		
Track Electrification (USSR)			1965	27
Replacement of Steam Locomotives	1950	12	1960	13

the growth of the total length of transport infrastructures, the time constants involved are very long indeed; it takes around eight decades to go from one percent to 50 percent of the ultimate final network length. This process, spanning over successive generations of diffusion life cycles of individual transport infrastructures and across major social and economic transformations, can therefore be considered as a good example of a truly large pervasive system.

At the next hierarchical level, when considering the growth of individual transport infrastructures, we observe that the building up of major infrastructures proceeds with a Δt of typically between four and five decades. As the next hierarchical level in the temporal diffusion process, we consider the case of technological change or upgrading *within* an already existing infrastructure grid, i.e., the substitution process of treated wooden railway cross-ties in the USA (a substitution process occurring in a contracting market, as the railway network of the USA has been decreasing since the 1930s) and the electrification of railway tracks in the case of the USSR. Here diffusion time constants (Δt) below three decades are typical, i.e., about twice as fast as in the case of new constructions of infrastructure grids.

Finally, at the last hierarchical level of the time constants of diffusion presented in Table 5.2, we consider a replacement process within the capital vintage structure in the form of the rolling stock of railways. The replacement of the coal fired steam locomotive by diesel traction (as in the USA) and by electric and diesel traction locomotives (in the USSR) proceeds in both countries with a Δt of 12 to 13 years, i.e. a similar diffusion constant. If we look at the hierarchical structure of temporal diffusion processes one level further to the micro level one observes even faster diffusion Δts, in the order of a few years for consumer durables. Even one hierarchical level further down, Δts of a few months can be observed, when considering the diffusion of fashion items or novelty gadgets.

Thus, we propose that diffusion processes can be characterized (classified) as operating within a temporal hierarchical structure, when considering the time constants (Δt) involved in diffusion. This hierarchy extends from the macro down to the micro level, and throughout we observe a self-similarity in the process of change, i.e., a fractal type of structure of diffusion processes all along their temporal (or spatial) hierarchical levels. Along the temporal hierarchy, diffusion processes appear to be faster the closer we get to the micro level. Of course more empirical studies will have to be carried out before such a working hypothesis can be firmly substantiated. However, we consider that a classification of diffusion processes along the hierarchy structure sketched out above may indeed be a first step in developing a quantitative taxonomy of diffusion processes.

Following the above taxonomy of diffusion processes and their temporal characteristics, we propose the following hypothesis as a weighting measure of different processes of change: A diffusion process is more important from a macro (economic) perspective, the higher it is located in the temporal hierarchy, i.e., the *longer* its diffusion time Δt. The proposed weighting measure thus links the *importance* of a particular diffusion process proportionally to its diffusion time constant Δt. Thus, we suggest that a one percent growth in the railway network of the USA (Δt of 54 years), is proportionally (54/12) more important than a one percent growth in the diffusion of diesel/electric locomotives (replacing steam locomotives) proceeding with a Δt of 12 years. As Figure 5.3 shows, the conclusion about the discontinuous nature of economic, technological, and social change in a dichotomy of accelerating and de-accelerating diffusion rates is also corroborated using this alternative measure. Figure 5.3 also presents results obtained by analyzing a larger sample of 265 innovation diffusion cases, revealing an identical pattern.*

Periods of accelerating rates of change indicate the emergence and growth of a new socio-technical paradigm in which a large number of technical, organizational, and social innovations diffuse into the economy and are absorbed by the society at large, contributing to prolonged periods of growth during which (energy) prices remain stable. These periods are followed by *seasons of saturations*, characterized by

---

* The generally higher absolute level of the average rate of change in the second (265 cases) data sample is not significant as it results from different weighting (by the respective mean Δt) in the two samples.

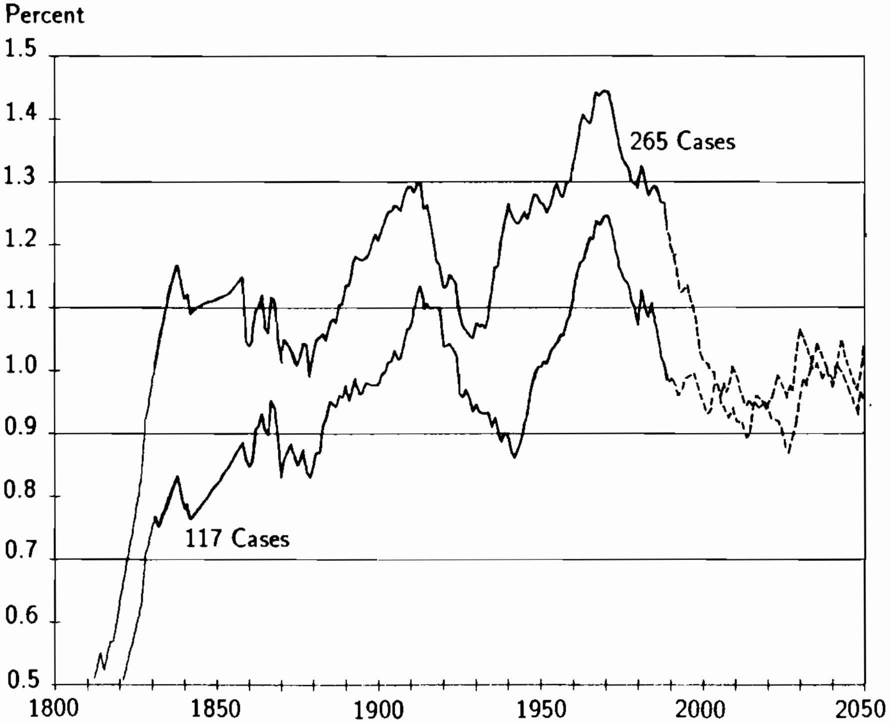


Figure 5.3. Weighted (by mean diffusion time) average rate of technical, economic, and social change in the USA, based on a sample of 117 and 265 innovation diffusion cases, in percent/year.

a significant decline in the average rate of technical and social change, accompanied by price flares indicating increased market volatility and turbulence. Schumpeterian "gales of creative destruction" (i.e., organizational readjustment processes like bankruptcies or intensive mergers), a lack of productive investments as a result of the saturation of traditional markets and uncertainty about the characteristics and investment opportunities of the new emerging paradigm prevail. Consequently we observe the widespread application of speculative instruments and increased market volatility in the form of "Black Fridays" or "Black Mondays" (recall here the October 1987 stock market crash). The shift in the search for profit opportunities away from productive investments is also reflected in massive investments into (speculative) mergers and acquisitions (over 300 billion US dollars in 1988 in the USA alone). From such a perspective, a redirection toward productive investments at a significant rate, may only finally occur once the opportunities opened up by an emerging socio-technical paradigm become more broadly perceived, and/or after large-scale destruction of speculative capital reduces the disparities in yields between financial and productive investments.

Seen from a broader international perspective, we may interpret the massive devaluation of currencies and capital stock (closures of unproductive nationalized industries) resulting from economic restructuring in most Eastern European countries and their "creative destruction" of social* and economic institutions during the second half of 1989, as being the most apparent manifestation of the downswing phase of the present Kondratieff cycle. In Western market economies the Kondratieff downswing has up to now been most visible in chronic unemployment, a general productivity slowdown, and saturation of some traditional markets, such as the steel industry, and the redirection to financial investments discussed above. However, it is exactly during such periods of turbulence and structural transition that the characteristics (technologies, new institutions) of the form and direction of a next expansion phase are being shaped. Diffusion of this new paradigm provides in the longer-term new investment and productivity gain possibilities in emerging new growth sectors and infrastructures.

Although the results presented above, as well as their tentative integration into a (holistic) model of discontinuities in the long-term economic development, are tentative and require further empirical and theoretical corroboration, they are encouraging. There are three reasons for this conclusion:

- Empirical evidence of the Schumpeterian bandwagon argument has been assembled.
- The narrow definition of innovations, being for a long time confined to the technological area, has been partly overcome and now includes variables of economic and social change also.
- The results confirm that even without a rigid clustering and synchronization in the appearance, growth or saturation phase of innovation diffusion, discontinuities in the long-term rate of change can be identified.

At this point, let us return to the discontinuities of transport infrastructure development proper. Figure 5.4 integrates the evolution of physical infrastructures with discontinuities in the evolution of energy prices and of the rate of technical, economic, and social change in the USA. Synchronization between the discontinuities in the development of physical transport infrastructures and discontinuities

---

* On the diffusion of democracy see for instance Modelski and Perry, 1989.

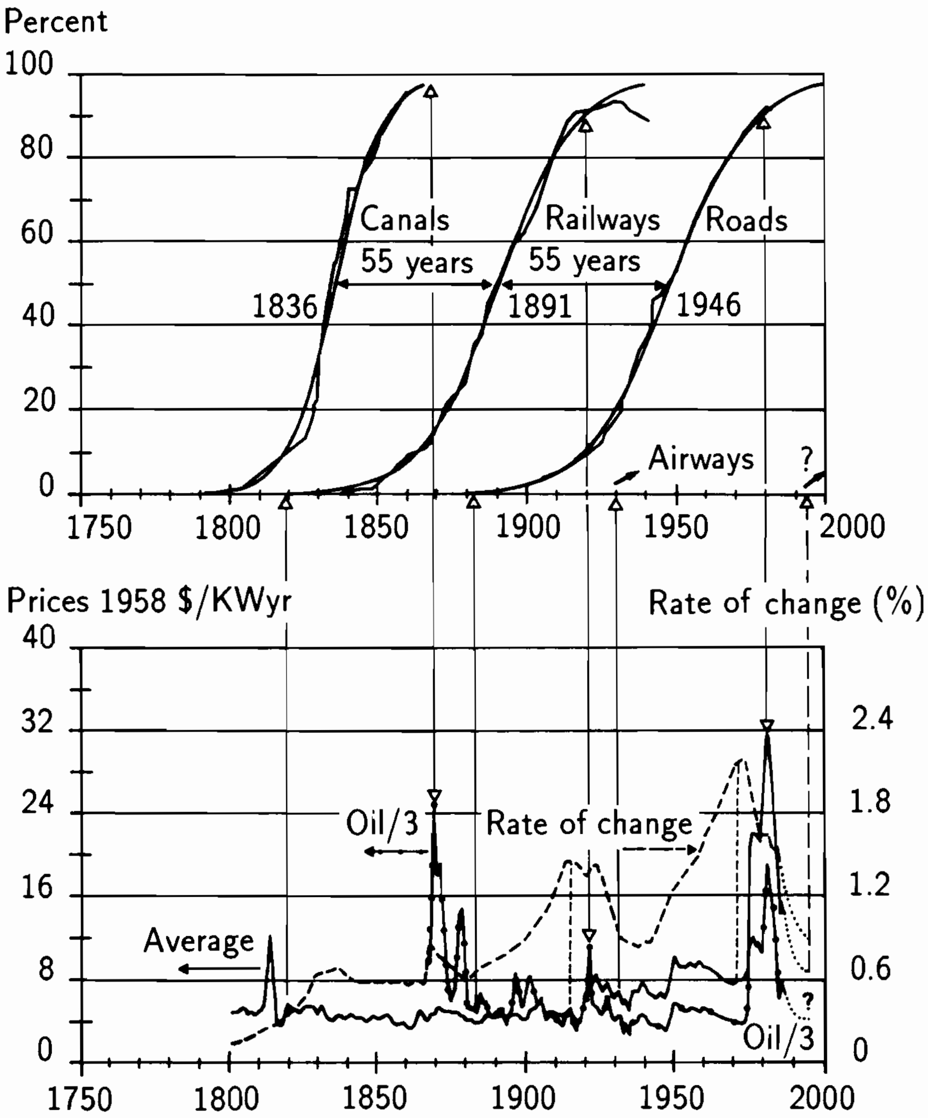


Figure 5.4. Discontinuities in the long-term evolution of average energy and crude oil prices (in 1958 US $/KWyr); the long-term rate of technical, economic, and social change (in percent/year); and growth of transport infrastructures (in percent of saturation length) in the USA.

in long-term economic development* emerges from Figure 5.4. The growth pulses of individual transport infrastructures are spaced at 55 year intervals (measured between the inflection points, i.e., the periods of maximum growth rates of canals, railways, and surfaced roads). The relationship between the periods of structural discontinuity in the evolution of transport infrastructures (saturation of growth of existing and emergence of new systems), the periods of price volatility and the turning points in the long-term rate of change are clearly discernible.

If economic history continues to unfold as it has in the past, the present (near-term) saturation of the expansion of the surfaced road network, as well as the possible emergence of a new transport infrastructure around the turn of the century would repeat past experiences. Based on such an assumption, we appear to be amid a period of structural discontinuity, characterized by the saturation in the expansion of an old development paradigm and the transition to a new, yet uncertain development path. In the past this structural discontinuity has been accompanied by high economic (price) volatility. The period of price volatility seems to have been almost overcome, and based on historical experience one could expect a return to stable real term energy and general price levels for an extended period of time (say 20 to 30 years).

The relationship of the dynamics of the development pattern of different infrastructures to long-term fluctuations in the economic development of the USA is not an isolated case. A similar chronological development pattern can also be shown in the evolution of different infrastructural bandwagons in a number of countries. Long waves are thus an international phenomenon. Figure 5.5 provides a concluding "summary of summaries" of the evolution of various transport technology clusters in the countries covered in the analysis in Chapter 3. The condensed results of the evolution of different transport systems discussed in Sections 3.1 to 3.4 above, are summarized to reveal the temporal overlap of their respective saturation phases.

The diffusion (growth) of different transport technologies and infrastructures in a number of countries is evolving within a relatively narrow band, defined by a diffusion frontier, i.e., the area spanned by the diffusion curves of the first and last country respectively, developing a particular transport system (canals, railways, roads and automobiles, air transport) or following a similar trajectory in the replacement of a particular transport technology (sail ships by steamships or horses by automobiles). We consider Figure 5.5 to be an empirical indication of the existence of international diffusion bandwagons and

---

* A similar argument was put forward by Hartman and Wheeler, 1979.

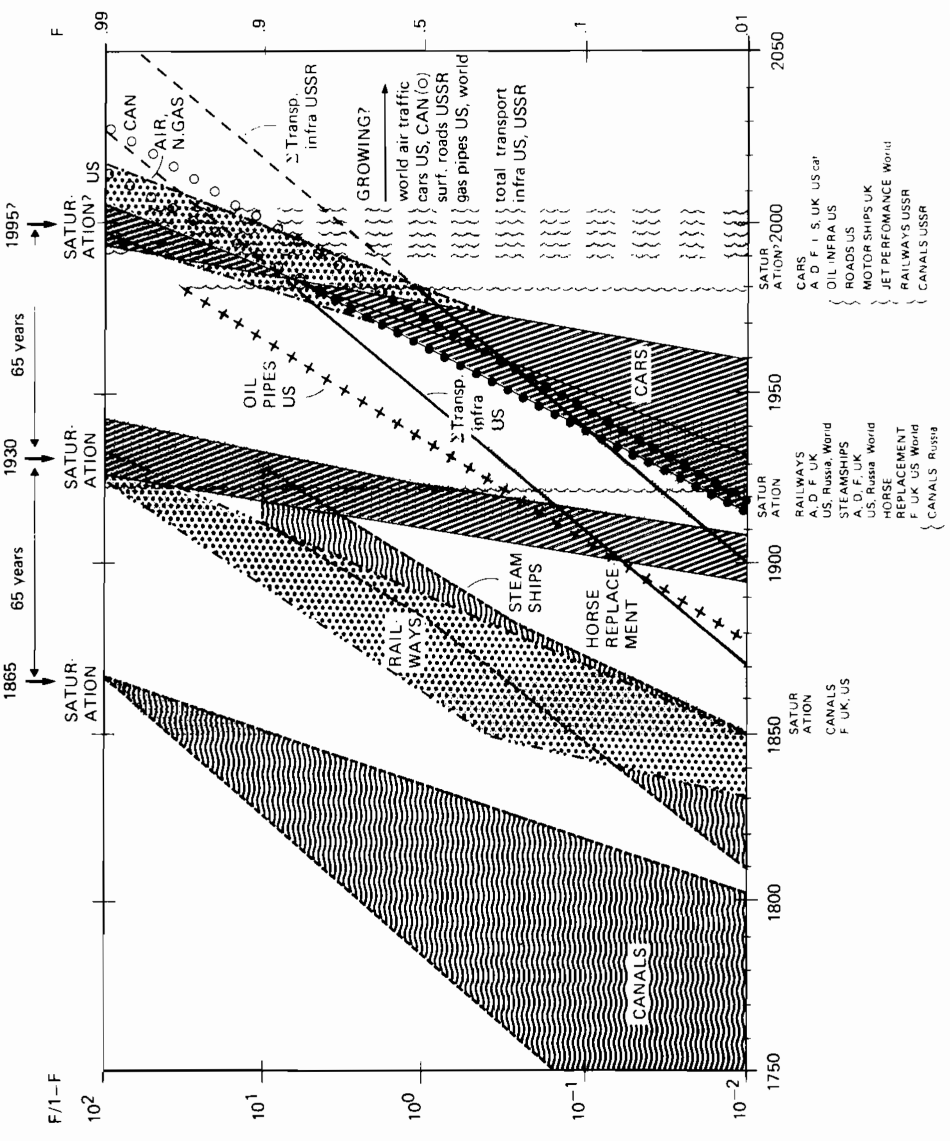


Figure 5.5. Diffusion and saturation in the development (growth) of canals, railways, steamships, horse replacement, automobiles and oil technology, and air transport and natural gas "bandwagons" in the main industrialized countries, and at the world level.

of the (converging) diffusion within technology clusters. As mentioned earlier, the focusing is not very pronounced and there is a high degree of overlap among the various diffusion bandwagons and technology clusters due to differences in the time constants (Δt) among individual diffusion processes. Nevertheless, three rather clear clusters can be distinguished. The first saturates around 1865, the second around 1930, and the third in 1995. Each cluster is converging toward the saturation period. The focusing increases as the diffusion cluster matures. In other words, the introduction of innovations is associated with great lags between the early and late adopters. However, the latecomers appear to achieve faster diffusion rates than the original innovators. We can say that there is a pronounced catch-up effect that converges toward the saturation point, indicating a *season of saturations* at the international level of technology diffusion as early as the 19th century.

The only noteworthy exception to this picture emerging from Figure 5.5 is the infrastructural development pattern of Tzarist Russia and the USSR. Whereas the development of the railway network and the replacement of sail ships in Tzarist Russia appear to be in tune with international developments, canal construction in Tzarist Russia enters its phase of saturation one cycle (around 50 years) later than in other countries. The transport infrastructure development policy of the USSR also falls outside the picture outlined in Figure 5.5. We have argued above, that this particular pattern may be an indication that the USSR has, to a large extent, been "decoupled" from the major economic expansion pulse that characterized the post World War II growth of Western market economies. It is thus not incidental that these past decoupling tendencies have become increasingly recognized within the USSR. In light of this, *perestroika* appears to be a possibly long overdue restructuring effort toward a new development path.

For all other countries individual transport development "bandwagons" are "rolling along in tune". Particularly noteworthy is the synchronization of the saturation periods inside the railway, steamship, and horse replacement "bandwagons" and of the three "bandwagons" together in the 1930s. We thus contend, that the severity of the Great Depression in the 1930s may indeed have been caused by the saturation of most of the technologies responsible for the economic upswing throughout the 19th century (i.e., the steam/coal technological cluster) and that this did not happen in individual countries alone, but at the international level.

Only a few technologies "tunneled through" (i.e., continued their growth pattern) during the 1930s depression, as illustrated by the example of oil pipeline growth in the USA in Figure 5.5. It was precisely this technological (i.e., oil) cluster, which in its growth trajectory was not noticeably affected by the saturation of the predominant technological (i.e., steam and coal) system, that became a key element in the subsequent growth period. Similarly, the growth trajectory of air transport and the development of natural gas infrastructures again appear to "tunnel through" the present or near-term saturation of the automobile age. This could be an indication that these technologies may become of pervasive importance in the future.

In this sense each Kondratieff long wave portrays a barrier to diffusion. Most processes saturate at the end of the inflationary period and the onset of the disinvestment phase in the Kondratieff wave. Very few diffusion processes can "tunnel" through this barrier. If it is true that this marks the beginning of paradigm shifts, it is not surprising that further diffusion of systems associated with the old techno-economic development trajectory is blocked to make way for new ones. It is the disruptive crisis of the old that provides the fertile ground for new systems to develop.

CHAPTER 6

Summary and Conclusion

The objective of this study is to give a quantitative long-term view of the history of individual transport infrastructures and to integrate these developments into a holistic picture of the evolution of the transport system as a whole. Innovations in transport and communication infrastructures and the technologies and devices using them, alter the range of human activities in space and time. Although generally technical in nature, they are closely related to important organizational and social adaptation processes. Some innovations become pervasive in the sense that they lead to a host of interrelated activities across many sectors of the economy or change many aspects of everyday life. Transport systems and infrastructures are examples of such pervasive innovations: they emerge, grow, and eventually decline as a result of changing societal requirements.

The main theme of this study is the dynamics of when and where innovations appear and become accepted by society, as manifest in their evolution (growth) and rate of interaction (replacement or substitution) with existing techniques. The regularity shown in the growth and structural changes of transport systems in different countries, even with different political systems and market clearance mechanisms, is consistent with the approach adopted in a new line of evolutionary economic models. In such models, ordered, structured evolutionary paths at the macro level are driven rather than dissipated by diversity in technological design, individual behavior, and the economic and social consequences of technological change at the micro level. Thus, the long-term development and structural changes in the transport sector portray features of self-organizing systems.

Our analysis has shown that the innovation diffusion and substitution models used within the present context for the systematization and organization of empirical data map the historical development patterns quite accurately. As such, these models constitute a necessary – and in their analytical resolution surprisingly accurate – tool for the quantitative description of long-term evolutionary processes. The use of quantitative diffusion and substitution models is also a prerequisite for international comparability, in order to take into account the varying boundary conditions of different systems, i.e., countries. Thus, developments in different countries can be compared in terms of their *rates of change* and in terms of the resulting *structural changes* of the transport system. Static cross-sectional comparisons of absolute diffusion levels appear from the perspective of this study to be highly problematic.

From a more formal viewpoint we conclude that, out of the numerous diffusion models proposed, the logistic growth and substitution model appeared as the most versatile and universally applicable in the cases analyzed here. This suggests that the underlying rationale for a symmetrical* diffusion model (i.e., a social learning process) is the basis for the diffusion patterns examined here. The speed at which diffusion/substitution proceeds, is a function of the "stimulus" (seen as a complex vector of comparative advantage changing over time, rather than a single static variable) for social acceptance of new infrastructures and for changing preferences underlying modal split decisions.

Growth and technological change in the transport sector follow regular, structured evolutionary paths, resulting in a characteristic pattern of change. This pattern consists of a sequence of replacements of older forms of transportation by new ones. We also showed that the basic elements of the technology life-cycle, birth, growth, saturation, and eventual decline, characterize the long-term development of individual transport technologies and infrastructures not only in different market economies, but also in planned economies. The pattern of temporal changes over time and between countries is marked by a high degree of regularity and a quest for higher speed and productivity.

This points to the fact that the underlying causal forces responsible for the historical development pattern transgress differences in political systems, market mechanisms, and relative cost/price structures. We maintain that the way in which different transport modes affect the spatial and temporal range of activities of a population determines the evolutionary development path observed. From such a

---

* This statement holds true only for the proper use of such diffusion models. Above all, an innovation must not be analyzed in isolation. As a result, diffusion can be modeled in most cases successfully only by accounting explicitly for technological *heterogeneity*, i.e., by use of a multiple technological substitution model.

perspective, it is primarily the quality and performance of different transport modes that influence modal split decisions. Under a constant travel time budget constraint, systems that are faster and allow for a wider human activity range gain a decisive comparative advantage over competitive systems.

Our analysis of the long-term trends in passenger travel suggest that as societies become more affluent, personal travel will grow significantly. People will also travel increasingly more for pleasure rather than for work. In addition we have concluded that, from a long-term perspective, transport and communication are of a complementary rather than substitutable nature. Under such premises, it appears again likely that we are in a phase of continuing long-term growth in passenger transport, especially as new information and communication technologies are on the verge of widespread diffusion. As regards goods transport, our analysis suggests that future changes in the transport system will most likely be driven by the evolving structure in output mix, further dematerialization, and a shift to higher value goods and services, as societies become wealthier. If, in fact, such tendencies become preeminent, economic growth will no longer be dependent on growing *volumes* of goods transport, but will impose higher quality standards for transport services.

The study suggests that out of the existing long-distance transport systems, air transport corresponds best to future requirements with respect to speed, flexibility and quality of service. Consequently, air traffic should be expected to grow in the future, including long-distance air freight. In the same way as urban centers grew in conjunction with the expansion of railways throughout the 19th century, airports could become new nodes of urban agglomerations, as they progressively evolve into centers of economic activity, e.g., as in the case of Frankfurt airport.

From the historical perspective, it appears that the success of any new transport system depends on a number of functional criteria. These include first, a synergistic development of a new technology (prime mover) base *in combination with* a new infrastructure system. Second, during the initial growth phase of a new infrastructure the *complementarity to*, and *independence of* existing infrastructures is important. Independence is possibly the most decisive factor in the emergence of new transport infrastructures because *new institutional and organizational* settings are required for the construction and successful operation of such systems.

Even when promising new transport systems appear to be the brink of commercial introduction, particularly in Japan and in a number of European countries, an over enthusiastic assessment of their short-term impact on the existing transport system is inappropriate. Older infrastructures are here to last, albeit with declining importance, as considerable time is required for building-up and decommissioning such large-scale systems. The life cycles spanning birth, growth and saturation and the start of decline of infrastructures are very long, lasting up to a century. The duration of senescence can be even longer. However, the most vital infrastructures are here to stay. Their immortality is marked by the provision of different services than were originally envisaged. More than a century after the canal age, the remaining inland waterways are used for leisure activities, irrigation, and maintaining a last market niche in the transport of low-value goods. There are also more boats today than in the heyday of ocean clippers, but they have entered a different market niche and are used as pleasure boats instead.

In their historical evolution, transport systems portray a strong homeostatic character in their growth and ordered structural transition paths, apparently being unaffected in the longer term by disruptions or major societal and economic discontinuities. Even in cases where major disruptive events resulted in a discontinuation of particular development trajectories, at least for a limited time period (e.g., in occupied France during World War II), the *rate* of evolution and structural transition, was not affected by such major discontinuities.

The analysis of the long-term development of different transport technologies, canals, railways, road and air transport, has shown that the historical evolutionary pattern may be described by:

- *Invariance* (similar development patterns can be observed across different countries and economic systems).
- *Regularity* (evolution via structured ordered growth and transition paths).
- *Discontinuity* (as a result of the emergence of new, and the saturation of old systems).
- *Synchronization* (empirical evidence of diffusion "bandwagons" at the international level as well as within given technological clusters was established).

We have observed that both technology clusters, like the transport technologies based on the steam engine, and country clusters, i.e., countries characterized by their similar diffusion pattern e.g., developing railway networks, tend to saturate during relatively short time periods. This "cluster" effect in the completion of a large number of interrelated technological diffusion and substitution processes and in a number of countries is referred to as a *season of saturations*. The technical and economic potentials for further growth of a particular technology cluster becomes exhausted, and awareness of increasing social disbenefits and environmental externalities from a further intensification of the dominant development mode become widespread, leading to structural change and transition to a next generation of transport systems.

This short time interval, during which most diffusion processes are completed and eventually enter a phase of decline is more noticeable in view of the long time constants involved in the growth of large-scale systems. The "simultaneous" saturation of infrastructure expansion pulses at the international level stems from the coupled dynamics of the introduction dates of systems and of the "catch-up" effect in the diffusion pattern between countries. Our understanding of the possible driving forces for the "season of saturations" is still insufficient to propose an underlying theoretical model. Here it is interpreted as being the result of the saturation and exhaustion of the growth potentials of the predominant paradigm of development, both nationally and internationally.

Our empirical results indicate that there is stronger evidence for a congruence in the saturation of diffusion processes rather than in their emergence, so that the focusing increases as systems mature. Such clustering is, however, not very rigid, but apparently still sufficient to result in major discontinuities in the long-term pattern of infrastructure development at the national and the international level. During the period of saturation of an old development trajectory, the previous generation of pervasive technologies, infrastructures, and institutional forms is not only challenged, but an opportunity window opens to replace the old paradigm in the longer term with new solutions. The overall development trajectory is thus punctuated by crises that emerge in the transition from an old, saturating to a new, yet uncertain development path.

Diffusion and replacement processes proceed according to a schedule that apparently defines the opportunity windows for the development of particular systems. In the "leader" countries diffusion leads to a long, sustained period of development with all the characteristics of pervasive systems, corresponding high adoption levels, and resulting in a specific spatial and social organization of societies. In "follower" countries where adoption occurs later, growth proceeds faster, but results in lower adoption levels and concomitant spatial and social structures. Consequently, railway networks are smaller in countries where they were introduced later; similarly, automobile ownership is also lower in countries catching-up on the leading countries. Since the evolutionary development path that follows saturation is based on fundamentally new techno-economic solutions, further development of the transport systems is based on new technologies and infrastructures rather than on a repetition of the growth characteristics of the previous development phase.* Thus, productivity gains from a further intensification of the traditional mode of development and its associated infrastructures are limited, leading to a transition to new systems. This even appears necessary for long-term development as Schumpeter (1935) observed: *"Add as many mail-coaches as you please, you will never get a railroad by so doing."*

A crucial question for future global resource and environmental impacts is whether Third World countries, many of them still embedded in the "First Wave" (Toffler, 1980), can proceed directly to a new, environmentally more benign "Third Wave" (Stigliani, 1990). To what extent will it be possible for developing countries to "leap-frog" entirely one technology life cycle such as massive private motorization? Historical experience suggests that while a particular technology system may not be skipped altogether, the intensity of its development may be orders of magnitude lower than the ones attained in a previous phase of economic development. Indeed, the study has illustrated cases (such as in the USSR) where the further intensification of traditional infrastructures appeared to impede rather than to promote successful catching up.

The integration of infrastructure growth and technological change in the transport sector into long-term economic development has shown that each major infrastructural development cycle is closely related to a cluster of important technological and organizational innovations. These form a "socio-technological paradigm" which dominates and drives major economic expansion periods, reaches saturation, and initiates a structural discontinuity in the long-term economic development paths.

The importance of the transport infrastructure developments appears crucial with respect to the observed pulses in economic activities. For example, the construction of canal networks throughout Europe and the USA during the 18th and the beginning of the 19th century was initiated by an increasing need to transport wood, charcoal, and other goods in large quantities over longer distances. Later, railroads, the steam engine, and coal-use were associated with a similar boom period basically due to the same reason: the concentration of production and population in urban areas required a more efficient transport system (and an energy carrier with higher energy density and versatility of use than traditional renewable energy sources). In a similar way, the growth of the automobile was contingent on the development of numerous other systems, such as paved roads, the internal combustion engine, oil refining and motor fuels, new sheet

---

* See Grübler and Nowotny, 1989, on this argument.

metals and high quality steels, electrical equipment and a whole host of other new technologies and products. Thus, we argue that specific expansion periods in economic development are driven by a cluster of interrelated technologies and infrastructures, which are interdependent and mutually cross-enhancing.

Transport systems in particular are prominent examples of pervasive technology clusters in various periods in history. Transport systems are however, closely linked and depend on the development of many other systems, particularly in the energy sector. Both the dominant transport and energy system approach maturity and saturation at the same time (either in the total length of the system or in terms of primary energy market shares): Canals and feed energy (for animal traction) around 1870, railways and coal around 1930, roads and automobiles together with oil, starting from the 1970s. This points to a common future growth potential for air transport and natural gas technologies and infrastructures, apart from their obvious interdependence for supersonic aircrafts (up to Mach 5 or 6), which could be fueled by methane. The "tunneling through" of air transport (and natural gas) through the present or near-term saturation of the diffusion of the automobile-oil "cluster", could be an indication of their growing importance in a forthcoming new expansion period in a similar way that railways and automobiles "tunneled-through" the saturation phase of their respective predecessor systems. Such a contention seems more plausible when considering the quality characteristics of these technologies.*

Future economic growth may proceed in a qualitative rather than quantitative manner, at least for developed economies. Increasing tendencies toward dematerialization, and higher information and software content (i.e., value) in industrial products, and in services, extensive new information and communication technologies, further growth in face-to-face communication (i.e., passenger transport), and increasing environmental compatibility through a closed industrial metabolism (Ayres, 1989b) etc., will all impose more stringent quality requirements on the transport and energy system.

If environment emerges as a major force determining the future direction of technological evolution, the successful development and introduction of new systems will only be possible if environmental boundary conditions and impacts, such as the effects of hypersonic aircrafts on the chemistry of the stratosphere, are endogenized right from the beginning into the design and systems configuration of new transport and energy systems. This is historically a new situation, but

---

* i.e., the comparative quality characteristics of air transport illustrated here, and the large resource base and environmental advantages of natural gas (see Lee *et al.*, 1988).

it enables new systems to become both environmentally and socially qualified, enhancing rather than inhibiting their growth prospects. Transport infrastructures are therefore closely associated with developments in the energy system and with the long-term evolution of the economy and of technological change in general.

Because of their vital importance to society, infrastructures and transport and energy technologies are, in fact "metaphor" (indicator) systems representative of historical economic expansion periods. The saturation in the expansion of a particular development paradigm leads to a phase of structural discontinuity and serves here as a model to explain slowdown phases in long-term economic growth. During this period of structural discontinuities and economic volatility, the decline of traditional, and the emergence of new systems is initiated. This "process of creative destruction" may serve as a theoretical model for the transitions and changes within the transport sector in the future, acting as an *essential and inherent part* of the long-term evolution of our economies and societies. It is the crisis of the old that provides the fertile ground for the emergence of the new.

Our transport infrastructure both records the past and shapes our present and future. It results from a long evolutionary sequence of growth and structural transition processes. The inherent dynamics of change in the long-term history of the transport systems imply that some technical and organizational/institutional solutions have entirely sunk into oblivion or are at the brink of disappearing. Feelings of regret about the loss of familiar infrastructure systems and the technologies using them are not warranted. Especially as these technologies do not *really* disappear, but rather "conquer" new market niches and frequently gain even a new, aesthetic value. The horse has disappeared as a means of transportation but is presently used for recreational purposes. Former canals, riverport and dock sites become preferred leisure and residential areas, railway lines are being turned into bicycle tracks. The Gare d'Orsay in Paris has been converted into a museum of the 19th century.

Will our future urban centers lie around airports, how will they be linked, and how will highway bridges be utilized, once they are no longer needed for automobiles? There remain far more questions than possible answers, but as Francis Bacon said: *He that will not apply new remedies must expect new evils; for time is the greatest innovator.*

References

Abernathy, W.J., 1978. The Productivity Dilemma, Roadblock to Innovation in the Automobile Industry. John Hopkins University Press, Baltimore and London.

Albertson, L.A., 1977. Telecommunications as a Travel Substitute: Some Psychological, Organizational, and Social Aspects. Journal of Communication, Vol.27, Nr.2:32-43.

d'Ancona, U., 1939. Der Kampf ums Dasein, eine biologisch-mathematische Darstellung der Lebensgemeinschaften und biologischen Gleichgewichte. Gebrüder Borntraeger, Berlin.

Angelucci, E. and Matricardi, P., 1977. Practical Guide to World Airplanes, Volumes 1 to 4., Mondatori, Milano (in Italian).

Annuaire Statistique (Ann. Stat.) de la France, 1961. Rétrospectif. Imprimerie Nationale, Paris.

Annuaire Statistique (Ann. Stat.) de la France, 1972 to 1987. INSEE, Paris.

Annuaire Statistique des Transports, 1982. Ministère des Transports, Département des Statistiques des Transports, Paris.

Abler, R., Adams, J.S. and Gould. P., 1971. Spatial Organization, the Geographer's View of the World. Prentice Hall, London.

Arcangeli, F., David, P.A. and Dosi, G. (Eds.), 1990. Frontiers on Innovation Diffusion. Oxford University Press (forthcoming).

Arrow, K.J., 1962. The Economic Implications of Learning by Doing. Review of Economic Studies, 29:155-173.

Arthur, W.B., 1983. Competing Technologies and Lock-in by Historical Small Events: The Dynamics of Choice under Increasing Returns. Center for Economic Policy Research Paper 43. Stanford University.

Arthur, W.B., 1988. Competing Technologies: An Overview, in: G. Dosi, *et al.* (Eds.), Technical Change and Economic Theory. Pinter Publishers, London and New York.

Ausubel, J.H., 1989. Regularities in Technological Development: An Environmental View, in: J.H. Ausubel and H.E. Sladovich (Eds.), Technology and Environment, US National Academy of Engineering, Washington D.C.

Ayres, R.U., 1969. Technological Forecasting and Long-range Planning. McGraw-Hill, New York.

Ayres, R.U., 1987. Future Trends in Factory Automation. WP-87-22, IIASA, Laxenburg, Austria.

Ayres, R.U., 1989a. Technology Forecast for CIM. Manufacturing Review, Vol.2, No.1:43-52.

Ayres, R.U., 1989b. Industrial Metabolism, in: J.H. Ausubel and H.E. Sladovich (Eds.), Technology and Environment, US National Academy of Engineering, Washington D.C.

BA Stat. (Bundesamt für Statistik), 1987. Statistisches Jahrbuch der Schweiz 1986, BA Stat. (Ed.). Birkhäuser Verlag, Basel.

Bandura, A., 1977. Social Learning Theory. Prentice-Hall, Englewood Cliffs, New Jersey.

Barker, T.C. and Robbins, M., 1975. A History of London Transport, Volume 1 and 2. Allen & Unwin, London.

Barry, B.J.L., 1988. Migration Reversals in Perspective: The Long-wave Evidence. International Regional Science Review Vol.11, No.3:245-251.

Bass, F.M., 1969. A New Product Growth Model for Consumer Durables. Management Science, 5:215-227.

Bass, F.M., 1980. The Relationship Between Diffusion Rates, Experience Curves and Demand Elasticities for Consumer Durable Technological Innovations, Journal of Business, 53:51-67.

Bass, F.M., 1986. The Adoption of a Marketing Model: Comments and Observations, in: V. Mahajan and Y. Wind (Eds.), Innovation Diffusion Models of New Product Acceptance. Ballinger, Cambridge, Mass.

Batten, D.F., 1989. The Future of Transport and Interface Communication: Debating the Scope for Substitution Growth, in: D.F. Batten and R. Thord (Eds.), Transportation for the Future. Springer Verlag, Berlin, Heidelberg, New York.

Bieshaar, H. and Kleinknecht, A., 1984. Kondratieff Long Waves in Aggregate Output? An Econometric Test. Konjunkturpolitik, Vol.30, No.5:279-303.

Blackman, A.W., 1971. The Rate of Innovation in the Commercial Aircraft Jet Engine Market. Technol. Forecasting and Soc. Change, 2:269-276.

Blackman, A.W., 1972. A Mathematical Model for Trend Forecasts. Technol. Forecasting and Soc. Change, 3:441-452.

Blackman, A.W., 1974. The Market Dynamics of Technological Substitutions. Technol. Forecasting and Soc. Change, 6:41-63.

Blackman, A.W., Seligman, E.J. and Sogliero, G.C., 1973. An Innovation Index Based on Factor Analysis. Technol. Forecasting and Soc. Change, 4:301-316.

Boeing, 1988. World Jet Airplane Inventory at Year End. Boeing Co., Seattle.

Borchard, K., 1968. Staatsverbrauch und öffentliche Investitionen in Deutschland 1780-1850. Dissertation an der Wirtschafts- und Sozialwissenschaftlichen Fakultät der Universität Göttingen.

Braudel, F., 1979. Civilisation matérielle, économie et capitalisme, XV^{e}-XVIII^{e} siècle. Le structure du quotidien: Le possible et l'impossible. Librairie Armand Colin, Paris.

O'Brien, P. (Ed.), 1983. Railways and the Economic Development of Western Europe 1830-1914. Macmillan Press, London.

Brooks, H., 1988. The Typology of Surprises in Technology, Institutions and Development, in: J. Clark and T. Munn (Eds.), Sustainable Development of the Biosphere. IIASA, Laxenburg, Austria.

Brown, L.A., 1981. Innovation Diffusion: A New Perspective. Methuen, New York.

Bruckmann, G., 1983. The Long-wave Debate. Options (1983)2:6-9, IIASA, Laxenburg, Austria.

Cameron, H.M. and Metcalfe, J.S., 1987. On the Economics of Technological Substitution. Technol. Forecasting and Soc. Change, 32:147-162.

Carlstein, T., 1982. Time Resources, Society and Ecology, Volume I: Preindustrial Societies. George Allen & Unwin, London.

Caron, F., 1983. France, in: P. O'Brien (Ed.), Railways and the Economic Development of Western Europe 1830-1914. Macmillan Press, London.

Carroll, G.R., 1982. National City-size Distributions: What Do We Know After 67 Years of Research? Progress in Human Geography, Vol.6, No.1 (March 1982):1-43.

Casetti, E., 1969. Why Do Diffusion Processes Conform to Logistic Trends? Geogr. Anal. 1.

Cerwenka, P., 1986. Wechselwirkungen zwischen Telekommunikation und Personenverkehr. Öffentlicher Sektor - Forschungsmemoranden, Vol.12, No.2 (June 1986):38-56.

Cerwenka, P. and Rommerskirchen, S., 1983. Aufbereitung globaler Verkehrsprognosen für die Fortschreibung der Bundesverkehrswegeplanung. Prognos AG., Basel.

Coleman, J., Menzel, H. and Katz, E., 1966. Medical Innovation: A Diffusion Study. Bobbs-Merrill, New York.

Daumas, M. (Ed.), 1980. A History of Technology and Invention, Progress Through the Ages, Volume III: The Expansion of Mechanization 1725-1860. J. Murray, London.

Daumas, M. and Gille, P., 1980. Part Three, Chapters 1 to 4, in: M. Daumas (Ed.), 1980. A History of Technology and Invention, Progress Through the Ages, Volume III: The Expansion of Mechanization 1725-1860, pp. 235-368. J. Murray, London.

Davies, S., 1979. The Diffusion of Process Innovations. Cambridge University Press, Cambridge.

Davies, S., 1980. Diffusion, Innovation and Market Structure, in: D. Sahal (Ed.), Research, Development and Technological Innovation, Recent Perspectives on Management. Lexington Books, Lexington, Mass.

De Green, K.B., 1988. Long Wave Cycles of Sociotechnical Change and Innovation: A Macropsychological Perspective. Journal of Occupational Psychology 61:7-23.

Debecker, A. and Modis, T., (Digital Equipment Corporation Geneva), 1986. Determination of the Uncertainties in S-curve Logistic Fits. Paper submitted to the Sixth International Symposium on Forecasting, Paris, June 15-18, 1986.

Delbeke, J., 1983. Recent Long Wave Theories, a Critical Survey, in: G. Bianchi, G. Bruckmann and T. Vasko (Eds.), Background Material for a Meeting on Long Waves, Depression and Innovation, Siena/Florence, October 26-29, 1983. CP-83-44, IIASA, Laxenburg, Austria.

DIW (Deutsches Institut für Wirtschaftsforschung), 1982. Analyse und Projektion der Personenverkehrsnachfrage in der Bundesrepublik Deutschland. Beiträge zur Strukturforschung Heft 70. Duncker & Humblot, Berlin.

DIW (Deutsches Institut für Wirtschaftsforschung), 1985. Regionale Struktur des Personenverkehrs in der Bundesrepublik Deutschland im Jahre 1980. Beiträge zur Strukturforschung Heft 88. Duncker & Humblot, Berlin.

Dixon, R., 1980. Hybrid Corn Revisited. Econometrica, 48:1451-1461.

Dosi, G., 1983. Technological Paradigms and Technological Trajectories. The Determinants and Directions of Technical Change and the Transformation of the Economy, in: C. Freeman (Ed.), Long Waves in the World Economy. Butterworths, London.

Dosi, G., Orsenigo, L. and Silverberg, G., 1986. Innovation, Diversity and Diffusion: A Self-organisation Model. Draft, SPRU, University of Sussex, June 1986.

Dosi, G., Freeman, C., Nelson, R., Silverberg, G. and Soete, L., 1988. Technical Change and Economic Theory. Pinter Publishers, London and New York.

DOT (Department of Transport), 1977. Transport Statistics Great Britain 1965-1975. Her Majesty's Stationary Office, London.

DOT (Department of Transport), 1987. Transport Statistics Great Britain 1976-1986. Her Majesty's Stationary Office, London.

Doxiadis C.A. and Papaioannou, J.C., 1983. Ecumenopolis. The Inevitable City of the Future. Athens Center for Ekistics, Athens.

Duckham, B.F., 1983. Canals and River Navigations, in: D.H. Aldcroft and M.J. Freeman (Eds.), Transport in the Industrial Revolution. Manchester University Press, Manchester.

van Duijn, J.J., 1983. The Long Wave in Economic Life. Allen & Unwin, London.

Dyos, H.J. and Aldcroft, D.H., 1969. British Transport, an Economic Survey from the Seventeenth Century to the Twentieth. Leicester University Press.

Eberlein, D., 1988. Projektionen des öffentlichen Personenvernverkehrs unter besonderer Berücksichtigung des grenzüberschreitenden europäischen Verkehrs. Thyssen Technische Berichte, Heft 1 (1988):39-50.

Der Elsner, 1985. Handbuch für Straßen und Verkehrswesen. Otto Elsner Verlagsanstalt, Darmstadt.

Easingwood, C.J., Mahajan, V. and Muller, E., 1981. A Nonsymmetric Responding Logistic Model for Technological Substitution. Technol. Forecasting and Soc. Change, 20:199-213.

Enzelberger, S., 1989. Ökonomische Hintergründe und Auswirkungen des langfristigen Technologiewandels im Verkehrssektor, dargestellt an den Beispielen USA, Großbritannien und Deutschland-BRD. Diplomarbeit, Institut für Finanzwissenschaft und Infrastrukturpolitik, Technische Universität Wien.

Feinstein, C.H., 1988. National Statistics 1760-1920, in: C.H. Feinstein and S. Pollard (Eds.), Studies in Capital Formation in the United Kingdom 1750-1920. Clarendon Press, Oxford.

Fischer Weltamanach, 1987. Fischer Taschenbuch Verlag, Frankfurt a.M.

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

Fishlow, A., 1965. American Railroads and the Transformation of the Antebellum Economy. Harvard University, Cambridge, Mass.

Floyd, A., 1968. A Methodology for Trend-forecasting of Figures of Merit. In: J. Bright (Ed.), Technological Forecasting for Industry and Government: Methods and Applications. Prentice-Hall, Englewood Cliffs, New Jersey.

Fogel, R.W., 1964. Railroads and American Economic Growth: Essays in Economic History. 1970 Edition, John Hopkins, Baltimore.

Forbes, R.J., 1975. Roads to c 1900, in: C. Singer, E.J. Holmyard, A.R. Hall and T.I. Williams (Eds.), 1975. A History of Technology, Vol. IV: The Industrial Revolution c 1750 to c 1850. Clarendon Press, Oxford.

Forrester, J.W., 1983. Innovation and Economic Change, in: C. Freeman (Ed.), Long Waves in the World Economy. Butterworths, London.

Fremdling, R., 1975. Eisenbahnen und deutsches Wirtschaftswachstum 1840-1879, ein Beitrag zur Entwicklungstheorie und zur Theorie der Infrastruktur. Gesellschaft für Westfälische Wirtschaftsgeschichte E.V., Dortmund.

Freeman, M.J., 1980. Road Transport in the English Industrial Revolution: An Interim Reassessment. Journal of Hist. Geography, Vol.6, No.1:17-28.

Freeman, C. (Ed.), 1983. Long Waves in the World Economy. Butterworths, London.

Freeman, C., Clark, J. and Soete, L., 1982. Unemployment and Technical Innovation: A Study of Long Waves and Economic Development. Frances Pinter, London.

Freeman, C. and Perez, C., 1988. Structural Crises of Adjustment, Business Cycles and Investment Behaviour, in: G. Dosi, *et al.* (Eds.), Technical Change and Economic Theory. Pinter Publishers, London and New York.

French, R.A., 1983. Canals in Pre-revolutionary Russia. Russian Historical Geography Vol.2:451-481.

Garrison, W.L., 1989. Using Technology to Improve Transportation Services, in: D.F. Batten and R. Thord (Eds.), Transportation for the Future. Springer Verlag, Berlin, Heidelberg, New York.

Gatignon, H.A. and Robertson, T.S., 1986. Intergration of Consumer Diffusion Theory and Diffusion Models: New Research Directions, in: V. Mahajan and Y. Wind (Eds.), Innovation Diffusion Models of New Product Acceptance. Ballinger, Cambridge, Mass.

Gayer, A.D., Rostow, W.W. and Schwartz, A.J., 1953. The Growth and Fluctuation of the British Economy 1790-1850, an Historical, Statistical and Theoretical Study of Britain's Economic Development. Clarendon Press, Oxford.

Gatto, M., 1985. Introduzione all'ecologia delle populazioni. CLUP, Milano.

Geise, J., 1959. What is a Railway? Technology and Culture, I:68-77.

General Motors, 1983. The First 75 Years of Transportation Products. Automobile Quaterly Inc., Princeton, New Jersey.

Ginarlis, J. and Pollard, S., 1988. Roads and Waterways 1750-1850, in: C.H. Feinstein and S. Pollard (Eds.), Studies in Capital Formation in the United Kingdom 1750-1920. Clarendon Press, Oxford.

Glaziev, S., 1988. Private communication, Central Economic-Mathematical Institute, Academy of Sciences of the USSR, Moscow. .

Godlund, S., 1952. Ein Innovationsverlauf in Europa, dargestellt in einer vorläufigen Untersuchung über die Ausbreitung der Eisenbahninnovation. Lund Studies in Geography, Ser. B. Human Geography No. 6., C.W.K. Gleerup Publishers, Lund, Sweden.

Goell, N.S., Maitra, S.C. and Montroll, E.W., 1971. On the Volterra and Other Nonlinear Models of Interacting Populations. Rev. Mod. Phys., 43:231-276.

Gold, B., Pierce, W. and Rosegger, G., 1970. Diffusion of Major Technological Innovations in the U.S. Iron and Steel Industry. Journal of Industrial Economics 18 (July 1970).

Goldstein, J.S., 1988. Long Cycles: Prosperity and War in the Modern Age. Yale University Press, New Haven and London.

Gompertz, B., 1825. Phil. Trans. Roy. Soc. London, 115:513-585.

Goodrich, C., 1960. Government Promotion of American Canals and Railroads 1800-1890. Columbia University Press, New York and London.

Goodrich, C., Rubin, J., Cranmer, H.J. and Segal, H.H., 1975. Canals and American Economic Development. Kennikat Press, Port Washington N.Y. and London.

Gordon, A., 1988. To Move With the Times, the Story of Transport and Travel in Scotland. The University Press, Aberdeen.

Gottinger, H-W., 1985. Modeling the Diffusion of Innovations. Internal paper, Batelle-Institut, Frankfurt, February 1985.

Green, A.G., 1986. Growth and Productivity Change in the Canadian Railway Sector 1871-1926, in: S.L. Engerman and R.E. Gallman (Eds.), Long-term Factors in American Economic Growth. University of Chicago Press, Chicago and London.

Grey, C.G. (Ed.), 1969. Janes's All the World's Aircraft, Reprint of 1919 Edition. David & Charles Publishers, London.

Griliches, Z., 1957. Hybrid Corn: An Exploration in the Economics of Technical Change. Econometrica, 25:501-522.

Grübler, A., 1988. Saturation Periods in Scientific-technical Progress. Paper delivered at the Conference on Regularities of Scientific-technical Progress and Long-term Tendencies of Economic Development, March 14-18, 1988, Novosibirsk, Akademgorodok.

Grübler, A. and Nakicenovic, N., 1987. The Dynamic Evolution of Methane Technologies. WP-87-2, IIASA, Laxenburg, Austria.

Grübler, A. and Nakicenovic, N., 1990. Long-waves, Technology Diffusion and Substitution. Review (forthcoming).

Grübler, A. and Nowotny, H., 1989. Towards the Fifth Kondratieff Upswing, Elements of an Emerging New Growth Phase and Possible Development Trajectories, in: Science, Technology and Development, International Council for Science Policy Studies, Paris.

Grübler, A., Posch, M. and Nakicenovic, N., 1988. Methods of Estimating S-shaped Growth Functions: Algorithms and Computer Programs. Internal paper, IIASA, Laxenburg, Austria.

Guile, B.R. and Quinn, J.B. (Eds.), 1988. Managing Innovation, Cases from the Services Industries. National Academy Press, 1988.

Hadfield, C., 1968. The Canal Age. David & Charles, Newton Abbot and London.

Hadfield, C., 1986. World Canals, Inland Navigation Past and Present. David & Charles, Newton Abbot and London.

Häfele, W. (Program Leader), 1981. Energy in a Finite World, a Global Systems Analysis. Ballinger, Cambridge, Mass.

Hägerstrand, T., 1952. The Propagation of Innovation Waves. Lund Studies in Geography, Ser. B, No. 4. Lund, Sweden.

Hägerstrand, T., 1953. Innovationsförloppet ur korologisk synpunkt (Engl. translation: Hägerstrand, 1967). University of Lund, Department of Geography Bulletin 15, Lund, Sweden.

Hägerstrand, T., 1967. Innovation Diffusion as a Spatial Process. University of Chicago Press, Chicago and London.

Hägerstrand, T., 1970. What to Do About People in Regional Science? Papers of the Regional Science Association XXIV, pp. 7-21.

Hamblin, R.L., Jacobsen, R.B. and J.L.L. Miller, 1973. A Mathematical Theory of Social Change. John Wiley & Sons, New York, London, Sydney, Toronto.

Hamilton, S.B., 1975. Building and Civil Engineering Construction, in: Singer, C., *et al.* (Eds.), 1975. A History of Technology, Vol. IV: The Industrial Revolution c 1750 to c 1850. Clarendon Press, Oxford.

Haritos, Z.J., 1987. The Economic Long Wave and Transportation. University of Toronto/York University, Toronto, Ontario.

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

Hawke, G.R., 1970. Railways and Economic Growth in England and Wales 1840-1870. Clarendon Press, Oxford.

Hefner, R.H. III, 1984. Natural Gas - the Politically and Environmentally Benign Least-cost Energy for Successful 21st Century Economies, the Energy Path to a Better World. GHK Company, Oklahoma City, Okla.

Heine, H., 1843. Lutetia, in: Allgemeine Zeitung, Augsburg 1840-1843. Reprinted in: Kindler Taschenbuch 1021-1024.

Herman, R. and Ausubel, J.H., 1988. Cities and Infrastructure: Synthesis and Perspectives, in: J.H. Ausubel and R. Herman (Eds.), Cities and Their Vital Systems, Infrastructure Past, Present and Future. National Academy Press, Washington, D.C.

Herman, R., Ardekani, S.A. and Ausubel, J.H., 1989. Dematerialization, in: J.H. Ausubel and H.E. Sladovich (Eds.), Technology and Environment, US National Academy of Engineering, Washington D.C.

Herodot, Historiae, Carolus Hude (Ed.) Oxford Classical Text, 3rd Edition 1927, Oxford University Press.

Historical Statistics of Canada, 1965, M.C. Urquhart and K.A.H. Buckley (Eds.). Cambridge University Press.

Hsia, R., 1973. Technological Change in the Industrial Growth of Hong Kong, in: B.R. Williams (Ed.), Science and Technology in Economic Growth. Macmillan, New York.

Hurter, A.P. and Rubinstein, A.H., 1978. Market Penetration by New Innovations: The Technical Literature. Technol. Forecasting and Soc. Change, 11:197-221.

IATA (International Air Transport Association), 1961 to 1988. World Air Transport Statistics. IATA, Cointrin-Geneva.

IRF (International Road Federation), 1970 to 1986. World Road Statistics. Geneva and Washington D.C.

Isard, W., 1942. A Neglected Cycle: The Transport Building Cycle. Rev. of Econ. Stat., XXIV, 4:149-158.

Jackman, W.T., 1962. The Development of Transportation in Modern England. Frank Cass & Co, London.

Janelle, D.G., 1969. Spatial Reorganization: A Model and Concept. Ann. Ass. Am. Geogr., LIX, 2:348-364.

Jane's, 1984. Jane's All the World's Aircraft 1984-1985, J.W.R. Taylor (Ed.), Jane's Publishing Co., London.

Jutila, S.T. and Jutila, J.M., 1986. Diffusion of Innovation in American Automobile Industry. Paper prepared for The Advanced Summer Institute in Regional Science, University of Umeå June 9-28, 1986. Umeå, Sweden.

Keith, C. and Grody, A., 1988. Electronic Automation at the New York Stock Exchange, in: B.R. Guile and J.B. Quinn (Eds.), Managing Innovation, Cases from the Services Industries. National Academy Press, Washington D.C.

Kleinknecht, A., 1987a. Innovation Patterns in Crisis and Prosperity, Schumpeter's Long Cycle Reconsidered. Macmillan Press, London.

Kleinknecht, A., 1987b. Are there Schumpeterian Waves of Innovations? WP-87-076. IIASA, Laxenburg, Austria.

Kommunist, 2, January 1987. Moscow. pp. 125-126.

Kondratieff, N.D., 1926. Die langen Wellen in der Konjunktur. Archiv für Sozialwissenschaft und Sozialpolitik, 56 (1926):573-609.

Kondratieff, N.D., 1928. Die Preisdynamik der industriellen und landwirtschaftlichen Waren. Archiv für Sozialwissenschaft und Sozialpolitik, 60 (1928):1-85.

Kruglikov, A.G., 1985. Dynamics Modeling of Technological Innovations Diffusion Processes. Paper presented at the Task Force Meeting on Dynamizing Technology October 28-29, 1985. IIASA, Laxenburg, Austria.

Laffut, M., 1983. Belgium, in: P. O'Brien (Ed.), Railways and the Economic Development of Western Europe 1830-1914. Macmillan Press, London.

Lee, T.H., Linden, H.R., Dreyfus, D.A. and Vasko, T. (Eds.), 1988. The Methane Age. Kluwer Academic Publishers, Dordrecht, Boston, London.

Lees, L.H. and Hohenberg, P.M., 1988. How Cities Grew in the Western World: A Systems Approach, in: J.H. Ausubel and R. Herman (Eds.), Cities and their Vital Systems, Infrastructure Past, Present and Future. National Academy Press, Washington, D.C.

Lepetit, B., 1983. Chemin te terre & voies d'eau, réseaux de transports et organisation de l'espace en France 1740-1840. Éditions de l'École des Hautes Études en Sciences Sociales, Paris.

Lewytzkyj, B. (Ed.), 1979. The Soviet Union, Figures - Facts - Data. K.G. Sauer, München, New York, London, Paris.

Linstone, H.A. and Sahal, D., 1976. Technological Substitution, Forecasting Techniques and Applications. Elsevier, New York, Oxford, Amsterdam.

Lord, C.L. and Lord, E.H., 1953. Historical Atlas of the United States, 1953 Edition. Henry Holt & Co., New York.

Lotka, A., 1924. Elements of Physical Biology. Republished 1956 as: Elements of Mathematical Biology, Dover Publications, New York.

Lufthansa, 1986. Start und Ziel: Berlin. Deutsche Lufthansa AG, Public Relations Direktion, Köln.

Mahajan, V. and Peterson, R.A., 1979. Integrating Time and Space in Technological Substitution Models. Technol. Forecasting and Soc. Change, 14:231-241.

Mahajan, V. and Peterson, R.A., 1985. Models for Innovation Diffusion. Sage Publications, Beverly Hills, London, New Dehli.

Mahajan, V. and Wind, Y., 1986. Innovation Diffusion Models of New Product Acceptance. Ballinger, Cambridge, Mass.

Mahajan, V., Muller, E. and Bass, F.M., 1990. New Product Diffusion Models in Marketing: A Review and Directions for Research. Journal of Marketing 54 (January 1990) (forthcoming).

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

Mansfield, E., 1961. Technical Change and the Rate of Imitation. Econometrica, Vol.29, No.4:741-766.

Mansfield, E., 1968a. Industrial Research and Technological Innovation, an Econometric Analysis. W.W. Norton & Co., New York.

Mansfield, E., 1968b. The Economics of Technological Change. W.W. Norton & Co., New York.

Mansfield, E., 1977. The Diffusion of Eight Major Industrial Innovations in the United States, in: N.E. Terleckyj (Ed.), The State of Science and Research: Some New Indicators. Westview Press, Boulder, Colorado.

Mansfield, E., 1984. Economic Effects of Research and Development, the Diffusion Process, and Public Policy, in: R.A. Goodman and Julián Pavón (Eds.), Planning for National Technology Policy. Praeger, New York.

Mansfield, E., Rapoport, J., Schnee, J., Wagner, S. and Hamburger, M., 1971. Research and Innovation in the Modern Corporation. Macmillan, New York.

Mansfield, E., Rapoport, J., Romeo, A., Villani, E., Wagner, S. and Husic, F., 1977. The Production and Application of New Industrial Technology. W.W. Norton & Co., New York.

Marchetti, C., 1975. On Strategies and Fate, in: W. Häfele, *et al.,* Second Status Report of the IIASA Project on Energy Systems 1975. RR-76-1, IIASA, Laxenburg, Austria.

Marchetti, C., 1980. Society as a Learning System: Discovery, Invention, and Innovation Cycles Revisited. Technol. Forecasting and Soc. Change, 18:267-282.

Marchetti, C., 1983. The Automobile in a Systems Context, the Past 80 Years and the Next 20 Years. Technol. Forecasting and Soc. Change, 23:3-23.

Marchetti, C., 1987. Infrastructures for Movement. Technol. Forecasting and Soc. Change, 32:373-393.

Marchetti, C., Nakicenovic, N., Peterka, V. and Fleck, F., 1978. The Dynamics of Energy Systems and the Logistic Substitution Model, Executive Summary. AR-78-1A, IIASA, Laxenburg, Austria.

Marchetti, C. and Nakicenovic, N., 1979. The Dynamics of Energy Systems and the Logistic Substitution Model. RR-79-13, IIASA, Laxenburg, Austria.

Marshall, C.F.D. 1938. A History of British Railways Down to the Year 1830. Oxford University Press, London.

Martino, J.P., 1983. Technological Forecasting for Decision Making. Elsevier, New York.

Martino, J.P., Chen, K-L. and Lenz, R.C., 1978. Predicting the Diffusion Rate of Industrial Innovations. UDRI-TR-78-42, University of Dayton Research Institute, Dayton, Ohio.

MBB (Messerschmitt-Bölkow-Blohm GmbH), 1989. Forecast to Year 2008. MBB Transport Aircraft Group, MBB, Hamburg.

Meadows, D.H., Meadows, D.L., Randers, J. and Behrens W.W., 1972. The Limits to Growth. Universe Books, New York.

Meißel, H., 1988. Private communication, Veterinary surgeon, Salzburg.

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

Metcalfe, J.S., 1970. Diffusion of Innovations in the Lancashire Textile Industry. Manchester School, June 1970.

Metcalfe, J.S., 1983. Impulse and Diffusion in the Study of Technical Change, in: C. Freeman (Ed.), Long Waves in the World Economy. Butterworths, London.

Metzer, J., 1974. Railroads in Tsarist Russia: Direct Gains and Implications. Explorations in Economic History, Vol.43, No.3:529-550.

Mitchell, B.R., 1980. European Historical Statistics 1750 - 1975, Second Revised Edition. Macmillan Press, London.

Mitchell, B.R., 1982. International Historical Statistics: Africa and Asia. Macmillan Press, London.

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

Modelski, G. and Perry, G., 1989. Democratization in Long Perspective, paper presented at the conference: Diffusion of Technologies and Social Behaviour, June 14-16 1989, IIASA, Laxenburg, Austria.

Montroll, E.W., 1978. Social Dynamics and the Quantifying of Social Forces. Proc. Natl. Acad. Sci. USA, Vol.75, No.10:4633-4637.

Montroll, E.W. and Badger, W.W., 1974. Introduction to Quantitative Aspects of Social Phenomena. Gordon and Breach, New York.

Morrill, R.L., 1968. Waves of Spatial Diffusion. J. Regional Sci., 8:1-18.

Morrill, R.L., 1970. The Spatial Organization of Society. Duxbury Press, Belmont, California.

Morrill, R.L., Gaile, G.L. and Thrall, G.I., 1988. Spatial Diffusion. Sage Publications, Beverly Hills, London, New Dehli.

Mosekilde, E. and Rasmussen, S., 1986. Technical Economic Succession and the Economic Long Wave. European Journal of Operational Research 25:27-38.

Mothes, F., 1950. Das Wachstum der Eisenbahnen. Zeitschrift für Ökonometrie, 1:85-104.

Munby, D.L., 1978. Inland Transport Statistics Great Britain 1900-1970, Volume I: Railways, Public Road Passenger Transport, London's Transport (edited and completed by A.H. Watson). Clarendon Press, Oxford.

Murillo, J.A., 1988. Modern Bridge Construction and Engineering Services, in: B.R. Guile and J.B. Quinn (Eds.), Managing Innovation, Cases from the Services Industries. National Academy Press, 1988.

MVMA (Motor Vehicle Manufacturers Association of the United States), 1983, and 1985. World Motor Vehicle Data. MVMA, Detroit and Washington D.C.

Myers, M.G., 1970. A Financial History of the United States. Columbia University Press, New York and London.

Nakicenovic, N., 1979. Software Package for the Logistic Substitution Model. RR-79-12, IIASA, Laxenburg, Austria.

Nakicenovic, N., 1984. Growth to Limits, Long Waves and the Dynamics of Technology. Dissertation an der Sozial- und Wirtschaftswissenschaftlichen Fakultät, Universität Wien.

Nakicenovic, N., 1986. The Automobile Road to Technological Change, Diffusion of the Automobile as a Process of Technological Substitution. Technol. Forecasting and Soc. Change, 29:309-340.

Nakicenovic, N., 1987. Dynamics of Change and Long Waves. Paper presented at the international workshop: Cycles de vie et cycles long: technologie, economie, société, July, 8-10, Montpellier, France. Reprinted as WP-88-074, June 1988, IIASA, Laxenburg, Austria.

Nakicenovic, N., 1988. Dynamics and Replacement of U.S. Transport Infrastructures, in: J.H. Ausubel and R. Herman (Eds.), Cities and their Vital Systems, Infrastructure Past, Present and Future. National Academy Press, Washington, D.C.

Narodnoye Khozyaistvo SSSR (Narod. Khoz.) 1922-1982, 1982. Finansi i Statistika, Moscow.

Narodnoye Khozyaistvo SSSR (Narod. Khoz.) 1986 and 1987. Finansi i Statistika, Moscow.

Nasbeth, L. and Ray, G.F. (Ed.), 1974. The Diffusion of New Industrial Processes: An International Study. Cambridge University Press, Cambridge.

National Resources Committee of the US Congress, 1937. Technological Trends and National Policy, Including the Social Implications of New Innovations, report of the Subcommittee on Technology (W.F. Ogburn, Chairman) to the National Resources Committee, 75th Congress, 1st Session, House Document No. 360, US Government Printing Office, Washington D.C.

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

Nelson, R.R. and Winter, S., 1982. An Evolutionary Theory of Economic Change. Harvard University Press, Cambridge, Mass.

Oster, S., 1982. The Diffusion of Innovation Among Steel Firms: The Basic Oxygen Furnace. Bell Journal of Economics, 13(1):45-56.

Österreichisches Statistisches Zentralamt (ÖStZA), 1975, 1980, and 1987. Statistisches Handbuch für die Republik Österreich. ÖStZA, Wien.

Pearl, R., 1925. The Biology of Population Growth. Knopf, New York.

Perez, C., 1983. Structural Change and the Assimilation of New Technologies in the Economic and Social System. Futures, Vol.15, No.4:357-375.

Peschel, M. and Mende, W., 1983, Leben wir in einer Volterra Welt? Akademie Verlag, Berlin.

Peterka, V., 1977. Macrodynamics of Technological Change: Market Penetration by New Technologies. RR-77-22, IIASA, Laxenburg, Austria.

Peterka, V. and Fleck, F., 1978. The Dynamics of Energy Systems and the Logistic Substitution Model, Volume 2: Theoretical Part. AR-78-1C, IIASA, Laxenburg, Austria.

Pollard, S., 1988. Coal Mining 1750-1850, in: C.H. Feinstein and S. Pollard (Eds.), Studies in Capital Formation in the United Kingdom 1750-1920. Clarendon Press, Oxford.

Pollins, H., 1971. Britain's Railways: An Industrial History. David & Charles, Newton Abbot.

Polk, R.L., 1987. Quoted in: M.C. Holcomb, S.D. Floyd and S.L. Cagle, 1987, Transportation Energy Data Book: Edition 9. Oak Ridge National Laboratory, Oak Ridge, Tenn.

Poor, H.V., 1860. History of the Railroads and Canals of the Unites States of America. John H. Schultz & Co, New York (reprinted by August M. Kelly Publishers, New York, 1970).

Pred, A., 1967. Postscript to English translation of T. Hägerstrand, Innovation Diffusion as a Spatial Process, pp. 299-324. University of Chicago Press, Chicago and London.

die Presse, September 9, 1988. U-Bahn-Raketen für den Reiseverkehr. ECO-Journal p. 5.

Prigogine, I., 1976. Order Through Fluctuation: Self-organization and Social System, in: E. Jantsch and C.H. Waddington (Eds.), Evolution and Consciousness. Addison-Wesley, Reading, Mass.

Pry, R.H., 1988. The Changing Nature of Industrial Management. Draft, IIASA, Laxenburg, Austria, March 1988.

Putzger, F.W. (revised by: Lendl, E., Wagner, W. and Klein R.), 1965. Historischer Weltatlas zur allgemeinen und österreichischen Geschichte, 44th edition. Hölder-Pichler-Tempsky, Österreichischer Bundesverlag, Wien.

Reinganum, J.F., 1983. Technology Adoption Under Imperfect Information. Bell Journal of Economics, 14(1):57-69.

Renouard, D., 1960. Les transports de marchandises par fer, route et eau depuis 1850. Librairie Armand Colin, Paris.

Robinson, J.P., Andreyenkov, V.G. and Patrushev, V.D., 1989. The Rhythm of Everyday Life, How Soviet and American Citizens Use Time. Westview Press, Boulder, Co., San Francisco, London.

Rogers, E., 1962. Diffusion of Innovations. The Free Press, New York.

Rogers, E. and Shoemaker, F., 1971. Communication of Innovations. The Free Press, New York.

Rogers, E., 1983. Diffusion of Innovations, 3rd Edition. The Free Press, New York.

Rogner, H-H., 1988. Technology and the Prospects for Natural Gas, Results of Current Studies. Energy Policy (February 1988):9-26.

Romeo, A., 1977. The Rate of Imitation of a Capital-embodied Process Innovation. Economica, Vol.44, No.173:63-69.

Rosenberg, N., 1976. On Technological Expectations. Economic Journal, 86:523-535.

Rosenberg, N. and Frischtak, C.R., 1983. Long Waves and Economic Growth: A Critical Appraisal. American Economic Review 73(1983):146-151.

Rostow, W.W., 1952. The Process of Economic Growth. W.W. Norton, New York, 2nd edition, 1960.

Rostow, W.W., 1975. Kondratieff, Schumpeter and Kuznets: Trend Periods Revisited. Journ. of Econ. Hist., 35:719-753.

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

Roszak, J. and Schmidt, B., 1986. Als das Dampfroß schnaufen lernte, die Eisenbahn in der Karikatur. Langen Müller, München, Wien.

Russell, T., 1980. Comments on 'The Relationship Between Diffusion Curves, Experience Curves, and Demand Elasticities for Consumer Durable Technological Innovations' (Bass, 1980). Journal of Business, 53:S69-S73.

Sahal, D., 1981. Patterns of Technological Innovation. Addison-Wesley, Reading, Mass.

Samaras, D.G., 1962. Nuclear Space Propulsion - a Historic Necessity. Nuclear Energy (September 1962):351-356.

Sammer, G., 1988. Private communication, Institut für Straßenbau und Verkehrswesen, Technische Universität Graz.

Sassin, W., 1981. The Coal Option, in: W. Häfele (Program Leader), 1981, Energy in a Finite World, a Global Systems Analysis. Ballinger, Cambridge, Mass.

Sax, E., 1920. Die Verkehrsmittel in Volks- und Staatswirtschaft, Volume 1 and 2. Second, revised Edition, Julius Springer Verlag, Berlin.

Schafer, E.H., 1971, Ancient China. Time-Life Books, Time Inc., New York.

Schipper, L., Barlett, S., Hawk, D. and Vine, E., 1989. Linking Life-styles and Energy Use: A Matter of Time? Annual Review of Energy, 14:273-320.

Schmidt, E., 1983. Quoted in Marchetti, 1983.

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

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

Schumpeter, J.A., 1942. Capitalism, Socialism and Democracy. 6th Edition, 1987, Unwin Paperbacks, London.

SETRA (Service d'Etudes Techniques des Routes et Autoroutes), 1988. Recensement de la circulation sur les routes nationales et autoroutes 1986. Ministère de l'Équipement, du Logement, de l'Aménagement du Territoire et des Transport, Paris.

Sharif, M.N. and Kabir, C., 1976a. A Generalized Model for Forecasting Technological Substitution. Technol. Forecasting and Soc. Change, 8:353-334.

Sharif, M.N. and Kabir, C., 1976b. System Dynamics Modeling for Forecasting Multilevel Technological Substitution. Technol. Forecasting and Soc. Change, 9:89-112.

Sharif, M.N. and Haq, A., 1979. Determinants for Forecasting Technological Substitution. Technol. Forecasting and Soc. Change, 13:59-81.

Sieferle, R.P., 1982. Der unterirdische Wald, Energiekrise und Industrielle Revolution. C.H. Beck, München.

Silvennoinen, P. and Väänänen, J., 1987. Forecasting Technological Substitution: The Logistic Model of Energy Systems Revisited. Technol. Forecasting and Soc. Change, 32:273-280.

Silverberg, G., 1984. Embodied Technical Progress in a Dynamic Economic Model, the Self-organizational Paradigm, in: R.M. Goodwin, M. Krüger and A. Vercelli (Eds.), Nonlinear Models of Fluctuating Growth. Springer Verlag, Berlin, Heidelberg, New York.

Silverberg, G., 1987. Technical Progress, Capital Accumulation, and Effective Demand: A Self-organizational Model, in: D. Batten, J. Casti and B. Johansson (Eds.), Economic Evolution and Structural Adjustment. Springer Verlag, Berlin, Heidelberg, New York.

Silverberg, G., Dosi, G. and Orsenigo, L., 1988. Innovation, Diversity and Diffusion: A Self-organisation Model. The Economic Journal 98 (December 1988):1032-1054.

Simon, H.A., 1988. Prediction and Prescription in System Modeling. Paper presented at the conference: IIASA: Perspectives and Future, 14-15 June, 1988, Laxenburg. IIASA, Laxenburg, Austria.

Simon, P., 1975. Models of Process Diffusion and Entry in the U.S. Chemical Industry. Ph.D. Diss. University of Pennsylvania.

Singer, C., Holmyard, E.J., Hall, A.R. and Williams, T.I. (Eds.), 1972. A History of Technology, Vol. II: The Mediterranean Civilizations and the Middle Ages c. 700 B.C. to c. A.D. 1500. Clarendon Press, Oxford.

Singer, C., Holmyard, E.J., Hall, A.R. and Williams, T.I. (Eds.), 1975. A History of Technology, Vol. IV: The Industrial Revolution c 1750 to c 1850. Clarendon Press, Oxford.

Singer, C., Holmyard, E.J., Hall, A.R. and Williams, T.I. (Eds.), 1979. A History of Technology, Vol. III: From the Renaissance to the Industrial Revolution c 1500 - c 1750. Clarendon Press, Oxford.

Skiadas, C., 1986. Innovation Diffusion Models Expressing Asymmetry and/or Positively or Negatively Influencing Factors. Technol. Forecasting and Soc. Change, 30:313-330.

SNCF (Direction des Études de la Planification et de la Recherche Département Statistique), 1987. Trafic de marchandises, débit journalier moyen. SNCF, Paris.

Snickars, F., 1987. The Transportation Sector in the Communications Society - Some Analytical Observations, in: P. Nijkamp and S. Reichmann (Eds.), Transportation Planning in a Changing World. Gower, Aldershot, Hampshire.

Sombart, W., 1916. Der moderne Kapitalismus, historisch-systematische Darstellung des gesamteuropäischen Wirtschaftslebens von seinen Anfängen bis zur Gegenwart, Volumes I to III, 2nd Edition. Dunker & Humblot, München und Leipzig (dtv reprint May 1987, DTV, München).

Sonis, M., 1984. Dynamic Choice of Alternatives, Innovation Diffusion, and Ecological Dynamics of the Lotka-Volterra Model, in: D.E. Pitfield (Ed.), Discrete Choice Models in Regional Science. London Papers in Regional Science Vol. 14. Pion Ltd., London.

Sonis, M., 1986. A Unified Theory of Innovation Diffusion, Dynamic Choice of Alternatives, Ecological Dynamics and Urban/regional Growth and Decline. Richerche Economiche, XL, 4:696-723.

Statistisches Bundesamt (Stat. BA.), 1979. Fachserie 7, Abschnitt 3: Generalhandel. Stat. BA., Wiesbaden.

Statistisches Bundesamt (Stat. BA.), 1987. Statistisches Jahrbuch 1987 für die Bundesrepublik Deutschland. Verlag W. Kohlhammer, Stuttgart, Mainz.

Statistisches Bundesamt (Stat. BA.), WiSta 7/1987, 1987. Einfuhr nach Grenzabschnitten und Verkehrszweigen 1986. Stat. BA., Wiesbaden.

Steinbach, J., 1980. Theoretische und methodische Grundlagen für ein Modell des sozialbestimmten räumlichen Verhaltens. Wiener Beiträge zur Regionalwissenschaft Band 3, Institut für Stadt- und Regionalforschung, Technische Universität Wien.

Stigliani, W.M., 1990. Private communication, International Institute for Applied Systems Analysis, Laxenburg, Austria.

Stürmer, G., 1872. Das Wachstum der Eisenbahnen, Entwicklung und jetzige Gestaltung sämtlicher Eisenbahnnetze der Erde. Mittler'sche Buchhandlung, Bromberg.

Szalai, A., 1972. The Use of Time, Daily Activities of Urban and Suburban Populations in Twelve Countries. Mouton, The Hague and Paris.

Tani, A., 1988. Penetration Mechanism of New Technologies: Integration Formulation from the Company Level to the Macroeconomic Level. WP-88-42, IIASA, Laxenburg, Austria.

Tarde, G., 1895, Les lois de l'imitation, étude sociologique, seconde édition. Germer Baillière et Cie., Paris

Teisserenc, 1845. Statistique des voies de communication en France. Paris, 1845 (data reproduced in Toutain, 1967).

Taylor, G.R., 1962. The Transportation Revolution, 1815-1860, Vol. IV of The Economic History of the United States. Hold, Rinehart and Winston. New York.

Thompson, F.M.L., 1976. Nineteenth-century Horse Sense. Econ. Hist. Rev. 29:67-72.

Toffler, A., 1980. The Third Wave. Morrow, New York.

Toutain, J.C., 1967. Les transports en France de 1830 à 1965. Cahiers de l'ISEA, Série Histoire Quantitative de l'Économie Francaise No. 8, Presses Universitaires de France, Paris.

Toutain, J.C., 1987. Le produit intérieur brut de la France de 1789 a 1982. Cahiers de l'ISMEA, Série Histoire Quantitative de l'Économie Francaise No. 15, Presses Universitaires de Grenoble, France.

UN, 1977. U.N. Yearbook of Industrial Statistics 1977, Volume 1: General Industrial Statistics. UN, New York.

UN, 1987. U.N. Industrial Statistics Yearbook 1987, Volume 1: General Industrial Statistics. UN, New York.

Urquhart, M.C., 1986. New Estimates of Gross National Product, Canada 1870-1926: Some Implications for Canadian Development, in: S. Engerman and R. Gallman (Eds.), Long-term Factors in American Economic Growth. University of Chicago Press, Chicago and London.

US DOC (U.S. Department of Commerce), 1975. Historical Statistics of the United States, Colonial Times to 1970, Volume 1 and 2. Bureau of the Census, Washington D.C.

US DOC (U.S. Department of Commerce), 1976 to 1989. Statistical Abstract of the United States 1976 to 1989, 97th to 109th Editions. Bureau of the Census, Washington D.C.

Van Gelderen (J. Fedder pseudonym), 1913. Springvloed: Beschouwingen over industrieele ontwikkeling en prijsbeweging. De Nieuwe Tijd 18.

Vasko, T. (Ed.), 1987. The Long-wave Debate. Springer Verlag, Berlin, Heidelberg.

Verhulst, P.-F., 1838. Notice sur la loi que la population suit dans son accroissement. Correspondence Mathématique et Physique 10: 113-121.

Verkehr in Zahlen, 1975, 1978, and 1987. Der Bundesminister für Verkehr (Ed.), Bonn.

Victor, D.G., 1989. Liquid Hydrogen Aircraft and the Greenhouse Effect. Working Paper, IIASA, Laxenburg, Austria (forthcoming).

Virilio, P., 1977. Vitesse et politique, essai de dromologie. Ed. Galilée, Paris.

Virilio, P., 1978. Fahren, fahren, fahren... Merve Verlag, Berlin.

Vivier, M., 1978. Transport des personnes et des marchandises. Paper presented at the conference: Role de l'Énergie Nucléaire dans le Cadre d'un Plan Energetique National. AIEA, Vienna, Austria and L'Institut National des Sciences et Techniques Nucléaires, Gif-sur-Yvette, France.

Voigt, F., 1965. Verkehr. Dunker & Humblot, Berlin.

Volterra, V., 1927. Una teoria matematica sulla lotta per l'esistenza. Scientia, Vol. XLI.

Volterra, V., 1931. Lecons sur la théorie mathématique de la lutte pour la vie (redigées par M. Brelot). Paris.

Voss, H., 1947. Die Entwicklung der Kraftfahrzeugstrassen in Deutschland. Dissertation an der Wirtschafts- und Sozialwissenschaftlichen Fakultät der Universität Köln.

Ward, W.H., 1967. The Sailing Ship Effect. Bull. Inst. Phys. & Phys. Soc. 18:169.

Wienecke, C., 1956. Entwicklungskritische Betrachtung des deutschen Straßenwesens in den Jahren 1871-1945. Forschungsarbeiten aus dem Straßenwesen, Neue Folge - Heft 28, Kirschbaum Verlag, Bielefeld.

Williams, R.H., Larson, E.D. and Ross, M.H., 1987. Materials, Affluence, and Industrial Energy Use. Annual Review of Energy, 12:99-144.

Woytinsky, 1927. Die Welt in Zahlen, fünftes Buch: Handel und Verkehr. Rudolf Mosse Buchverlag, Berlin.

Zahavi, Y., 1979. The "UMOT" Project. Report DOT-RSPA-DPB-20-79-3, prepared for U.S. Department of Transportation and Ministry of Transport, FRG. NTIS, Springfield, Va.

Zahavi, Y., 1981. The UMOT-urban Interactions. DOT-RSPA-DPB 10/7, U.S. Department of Transportation, Washington, D.C.

Zipf, G.K., 1949. Human Behavior and the Principle of Least Effort. Hafner Publishing Co., New York, 1972 (reprint of 1949 edition).

Zuzanek, J., 1980. Work and Leisure in the Soviet Union, a Time-budget Analysis, Praeger, New York.

Contributions to Economics

The series "Contributions to Economics" contains publications in the fields of economics and management science. These publications are primarily monographs and multiple author works containing new research results, but conference and congress reports are also considered.

Apart from the contribution to scientific progress presented, it is a notable characteristic of the series that actual publishing time is very short thus permitting authors and editors to present their results without delay.

Manuscripts could be sent directly to Physica-Verlag:

Physica-Verlag GmbH & Co.
Tiergartenstraße 17
6900 Heidelberg, FRG

ISBN 3-7908-0479-7
ISBN 0-387-91374-2