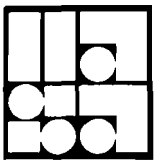


The Dynamics of Energy Systems and the Logistic Substitution Model

Volume 2: Theoretical Part

V. Peterka
F. Fleck

Prepared for
the Stiftung Volkswagenwerk

Administrative Report
AR-78-1C
July 1978

Energy Systems Program

International Institute for Applied Systems Analysis
A-2361 Laxenburg, Austria Telephone: 02236/7521'0

THE DYNAMICS OF ENERGY SYSTEMS AND THE LOGISTIC SUBSTITUTION MODEL

Volume 2:

THEORETICAL PART

V. Peterka
F. Fleck

July 1978

INTRODUCTION

The application of a certain rule, even if always successful in practice, confronts very strong psychological barriers if it lacks a certain logical frame linking it causally to the body of accepted knowledge.

Phenomena rooted in social behavior are always very difficult to "explain" in such a way, because if we rely too much on the basic, irrational, and stochastic roots of our decisions, then the explanation is rejected as "too mechanistic".

If on the other hand we rely on the perception people have of themselves, as rational and wise decision makers, then we fall into a maze of ad hoc explanations that strongly resembles local politics.

Economists, who have faced a much similar problem, have made a great, partially successful, effort in describing and organizing the monetary measurables of human activity. Although they too miss primary causes they can introduce concepts of minimization and optimization which permit choices and structuring of the systems.

Our attempt to "hook" the market penetration rules to the accepted scientific system have followed both routes.

Fleck takes the stochastic "irrational" view. Social processes, and introduction of a new technology is a social process, are seen as the envelope of a maze of tiny decisions, causally unrelated, and, like nails in the path of a falling ball, slowing down its chute and "diffusing" its landing point. A good social example of this process is given by the diffusion of an infection e.g. the common flue. Although in a case-by-case analysis the biologist can give a fair causal description of the process, the contacts that lead to the diffusion are within another realm of causality and are better described stochastically.

Learning processes are well described in such a way, and they yield logistic curves. Fleck then visualizes the diffusion of a technology as a social learning process under constraints. The stability of the curves is a mark of the stability of man and society as learning structures.

The weak point of the theory is that the critical parameters have to be measured post hoc, and they are not reducible to other measurements that could be made before the penetration is initiated.

Peterka on the other hand follows a more classical route, taking economics as a driving force. He assumes that an industry to expand has to generate profits. External capital can produce some time shifts, providing actual money for expected gains, but the picture is not much blurred.

Consequently, as substitution is driven by differential growth rates, these rates must be driven by differential profits. Perhaps a weak point of this theory is that differential profits must be constant (if smoothed) over long periods in order to produce well-behaved logistics. This feeds back to regular progress curves and automatic price leveling.

We can invert the reasoning and look for the stable progress curves and price leveling whose existence can be postulated from the very regular evolution of market penetration curves. This would greatly add to our understanding of the system.

The treatment by Peterka is quite general and produces curves which can specialize as logistic, but may also have more complex expressions. In general, the "graininess" of the data does not permit to distinguish between the various curves, and we usually stick to our logistic, which has the great advantage of straightforward simplicity.

Altogether we think that the basic objective of the grant has been fulfilled. We explored the field experimentally showing the great efficiency of our model in organizing data, and we tried two ways to bring its working under logical scrutiny.

The fact that during this operation we have presumably generated more problems than we solved is a good indication that we are plowing a fertile field.

CONTENTS

Macrodynamics of Technological Change: Market Penetration by New Technologies

V. Peterka

1. INTRODUCTION

Abstract mathematical models appear to be powerful tools in forecasting the future. It seems that there are two main reasons for their growing popularity and for growing endeavor to build mathematical models for more and more complex processes.

First, mathematical modeling makes it possible to decompose human reasoning into simpler steps and to express it in quantitative terms. In this way mathematical models help the forecaster to be objective and to avoid unintentional bias due to his natural efforts to make the future what he wants it to be.

Second, once a model is found and verified it gives a better insight into and a better understanding of the process studied, and, what perhaps is most important, also shows how the future development of the process can be influenced and controlled.

However, any mathematical model and any mathematical theory can be only a simplified image of the objective reality or of the laws by which the reality is governed. A very detailed and thorough model may even be undesirable if it is too complicated and difficult to apply. A good mathematical model should have the following properties.

- It should reflect the relations that are most important for the purpose for which the model is built. The difficulty is that usually it is not a priori clear which relations are important and which can be neglected. Therefore the development of a model is, as a rule, an iterative and learning procedure.

- It should be as simple as possible. By simplicity is meant here, first of all, the low number of parameters

that have to be determined. Mathematics supplies the model builder with an immense number of possibilities for describing a particular relation. However, only a correct choice of the structure of the model makes it possible to reduce the number of parameters by which all possible cases can be characterized and to minimize the number of exogenous quantities and variables. The choice of the model structure is perhaps the most critical step in model building.

- It should be based on assumptions that are well understood. As any model can be only an approximate description of the complex reality, it is true that assumptions are made to be violated. However, the simpler and clearer the assumptions are, the better the judgment that can be made about the reliability of the answers the model can give to our questions.

In general it may be very difficult to meet all the requirements formulated above and there is no unique way how to proceed optimally (if any optimum exists at all). Model building is and always will remain an art and a game: an art in how to combine mathematics, intuition, sound reasoning, and experience (one's own and of predecessors); a game between the human intellect and nature based on the rule of trial and error. This is what makes the mathematical modeling of the real world so attractive and exciting.

This paper deals with mathematical modeling of the dynamics of interaction between society and new technologies. The progress in technology can be viewed as a continuing historical process during which existing forms of satisfaction of human needs are replaced by new and superior ones.

Reliable forecasting of technological changes is surely of great interest for corporations and producers planning their activity and looking for new opportunities. However, it seems that the understanding of the diffusion of new and emerging technologies may be of much broader importance. Considering that man has few basic material needs to be satisfied--food, clothing, shelter, defense, transportation, communication, health care, and

entertainment--one can regard the material development of society as a sequence of substitution processes, or as a single multivariate substitution process with many technologies sequentially entering and leaving the process.

The impulse for this study was given by the work of Marchetti [1,2] who also oriented the author's attention toward the empirical model advanced by Fisher and Pry [3,4] for the case of two competing technologies. Inspite of a very low number of parameters, the curves generated by these models fit the known historical data with a precision which is much higher than one is used to in the modeling of economic and social systems. The main objectives of this study are:

- To explain the existing empirical models and rules which appear to fit the historical data, and to define the conditions under which they hold.

- To find a law governing multivariate substitution processes, i.e., with any number of competitors.

- To develop algorithms and computer programs for forecasting of substitution processes and for estimation of model parameters from historical data.

- To find formulae making it possible to calculate the model parameters on the basis of the economic assessment of a new technology in order to be able to incorporate the newcomers also in the model.

- To develop a probabilistic model of the substitution process, making it possible to describe and evaluate the accuracy of forecasting.

- To apply the model to substitution of primary energy sources in world energy consumption, to verify the model on historical data (wood, coal, oil, natural gas), and to show how the possible role of a new energy source (e.g. nuclear) can be forecast on the basis of objective and quantitatively well defined data.

The paper is organized in the following way.

In Section 2 existing empirical models of substitution processes are briefly reviewed and discussed. Several examples are given to illustrate the problem and to point out some important facts.

The approach adopted in this study is based on a differential equation formulated in Section 3. This equation reflects the long-term balance of capital flows governing the production using one of competing technologies. A system of such equations describes the multivariate competition and the introduction of market shares makes it possible to eliminate the market price. Following this basic idea a simple model is derived in Section 4, which can be considered as an extension of the Fisher-Pry model to the multivariate case. This model is derived under certain simplifying assumptions; they are removed in Section 5, where a more general model is presented. As a closed analytical solution does not exist in this case an algorithm is developed which solves the problem numerically in a very effective way.

The question how the information about the model parameters can be extracted from known historical data is studied in Section 6. To be able to answer this question in a consistent way, it was necessary to extend the deterministic models developed in previous sections to a probabilistic model. The problem of parameter estimation is solved in detail including the numerical algorithms and practical examples.

The problem of forecasting is addressed in Section 7. It is shown how the uncertainty of model parameters can be respected in forecasting and how a new technology can be incorporated in the model using its economical assessment. This is demonstrated on the forecast of market penetration by nuclear energy.

In the concluding Section 8 the range of validity of the model is discussed from a general point of view and some further possible applications are outlined.

Technical details of mathematical character, which have been omitted in the main text to make it easier to follow, can be found in Appendix A in the form of mathematical theorems and proofs.

In Appendix B a simple and effective optimization method is developed. It is applied in maximum likelihood estimation in Section 6.

The main practical results of the theory presented are condensed in FORTRAN - subroutines the commented listings of which can be found in Appendix C.

2. FISHER-PRY EMPIRICAL MODEL

Perhaps the first systematic attempt at forecasting technological changes based on a mathematical model is due to Fisher and Pry [3,4]. They collected historical data on a wide variety of substitutions and advanced a model which fits existing data remarkably well. The results of Fisher and Pry's investigation apply to two competing technologies of commodities fulfilling the same need. The essence of their work can be stated as follows.

Let $f_1(t)$ be the fraction of market occupied by the commodity produced by the first, old technology at time t and $f_2(t)$ the fraction of market penetration by the second, new technology at the same time.

$$f_1(t) + f_2(t) = 1 \quad . \tag{2.1}$$

If the observed values of f_2 are plotted as a function of time the plot follows with a high regularity the S-shaped curve given in Figure 1.

This curve can be described by the following formula:

$$f_2(t) = \frac{1}{2}\left[1 + \operatorname{tgh} \frac{c_{12}}{2} (t - t_h)\right] = \frac{1}{1 + e^{-c_{12}(t - t_h)}} \quad , \tag{2.2}$$

where t_h is the time at which the half of the market is penetrated, $f_2(t_h) = f_1(t_h) = \frac{1}{2}$. For given t_h the entire substitution process is determined by a single parameter c_{12} which is denoted in the

original paper as 2α; we changed the notation to be compatible with the rest of the paper where more than two competing technologies are considered.

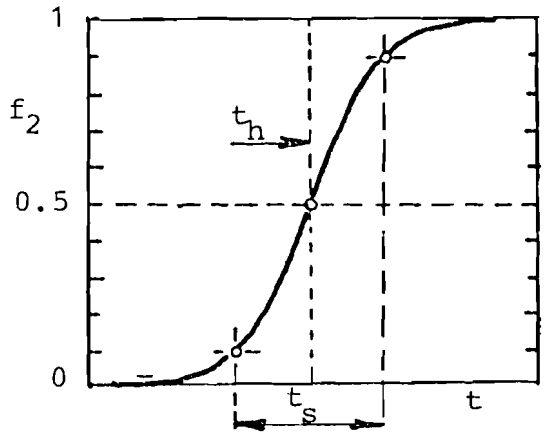


Figure 1. Logistic curve.

The fraction $f_1(t)$ can be obtained simply as a supplement to one according to (2.1), or formally from the formula (2.2) when the indices 1 and 2 are interchanged and the relation

$$c_{21} = -c_{12} \tag{2.3}$$

is used.

The relation (2.2) can be rearranged in the following way:

$$\frac{f_2(t)}{1 - f_2(t)} = e^{c_{12}(t - t_h)} \tag{2.4}$$

or

$$\ln \frac{f_2(t)}{1 - f_2(t)} = c_{12}(t - t_h) \quad . \tag{2.5}$$

This indicates that the substitution data, when plotted in the form of $f_2/(1 - f_2)$ as a function of time semilogarithmically, should form a straight line. This appeared to hold with extraordinary precision for a wide range of cases investigated.

Fisher and Pry defined the "takeover time" of the substitution as the time period t_s required to transfer from $f_2 = 0.1$ to $f_2 = 0.9$. It is easy to find that the takeover time t_s and the parameter c_{21} are related in the following way:

$$c_{12} = \frac{2 \ln 9}{t_s} \quad .$$

If the dimensionless time $\tau = \frac{2(t - t_h)}{t_s}$ is introduced, formulae (2.2) and (2.5) can be written in the following parameterless forms:

$$f_2(\tau) = \frac{1}{1 + 9^{\tau}} \quad ; \quad \ln \frac{f_2(\tau)}{1 - f_2(\tau)} = \ln 9 \cdot \tau \quad .$$

This makes it possible to plot different substitution processes into a single graph. Fisher and Pry have done it for 17 substitutions listed in Table 1. The result is shown in Figure 2. For more details the reader is referred to [3]. Here, we will point out only that the main outliers in Figure 2 concern the synthetic/natural rubber substitution and are due to the perturbation that occurred during the Second World War when large effort was undertaken to support the substitution. This is clearly seen from Figure 3 [3].

A very important feature of the Fisher-Pry model is that it describes the evolution of the fractional market share and not the total production of the particular commodity. While the total production may be influenced by various and often unknown external factors the evolution of the fractional market share exhibits nice regularity. This can be clearly seen from the examples in Figures 4 and 5. The line for steel production in the USSR indicates that the model may be valid also for societies with planned economies.

For later use some other possible forms of the Fisher-Pry substitution model will now be given.

Table 1.

Substitution	t_s Years	t_h Year
Synthetic/Natural Rubber	58	1956
Synthetic/Natural Fibers	58	1969
Plastic/Natural Leather	57	1957
Margarine/Natural Butter	56	1957
Electric Arc/Open Hearth Specialty Steels	47	1947
Water Based/Oil-Based House Paint	43	1967
Open Hearth/Bessemer Steel	42	1907
Sulfate/Tree-Tapped Turpentine	42	1959
TiO_2/PbO-ZnO Paint Pigments	26	1949
Plastic/Hardwood Residence Floors	25	1966
Plastic/Other Pleasure Boat Hulls	20	1966
Organic/Inorganic Insecticides	19	1946
Synthetic/Natural Tire Fibers	17.5	1948
Plastics/Metal Cars	16	1981
BOF/Open Hearth Steels	10.5	1968
Detergent/Natural Soap (US)	8.75	1951
Detergent/Natural Soap (Japan)	8.25	1962

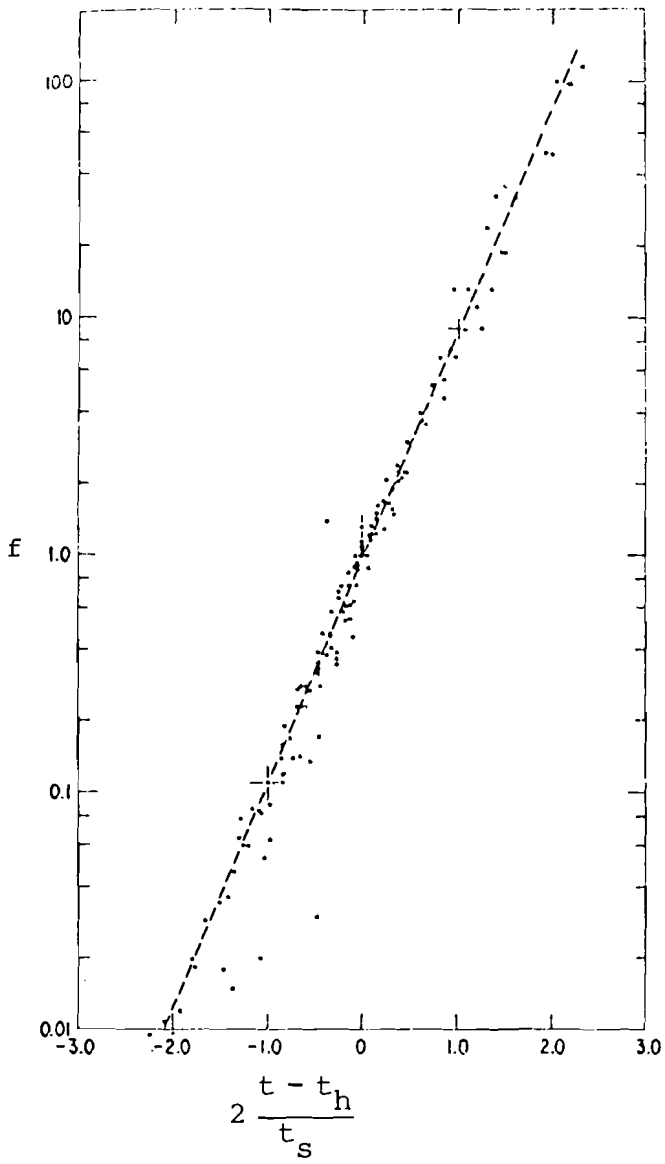


Figure 2a.

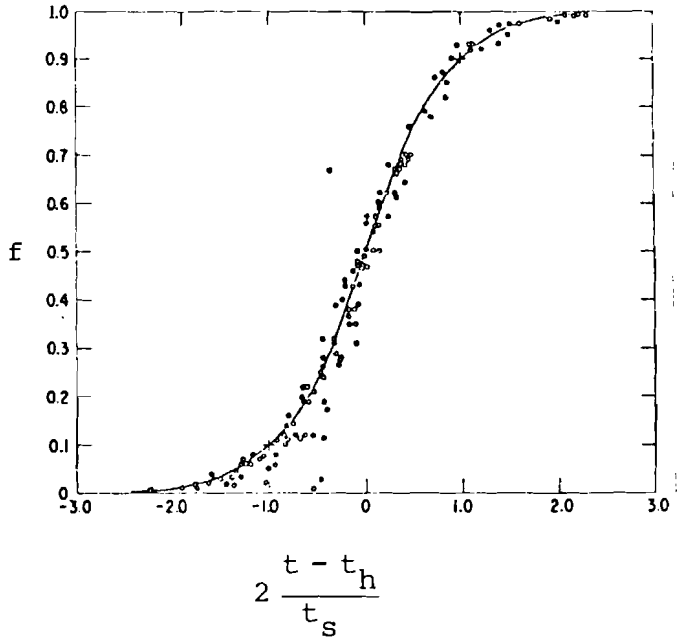


Figure 2b.

Figure 2. Historical data of 17 substitution processes plotted versus normalized units of time.

Source: [3]

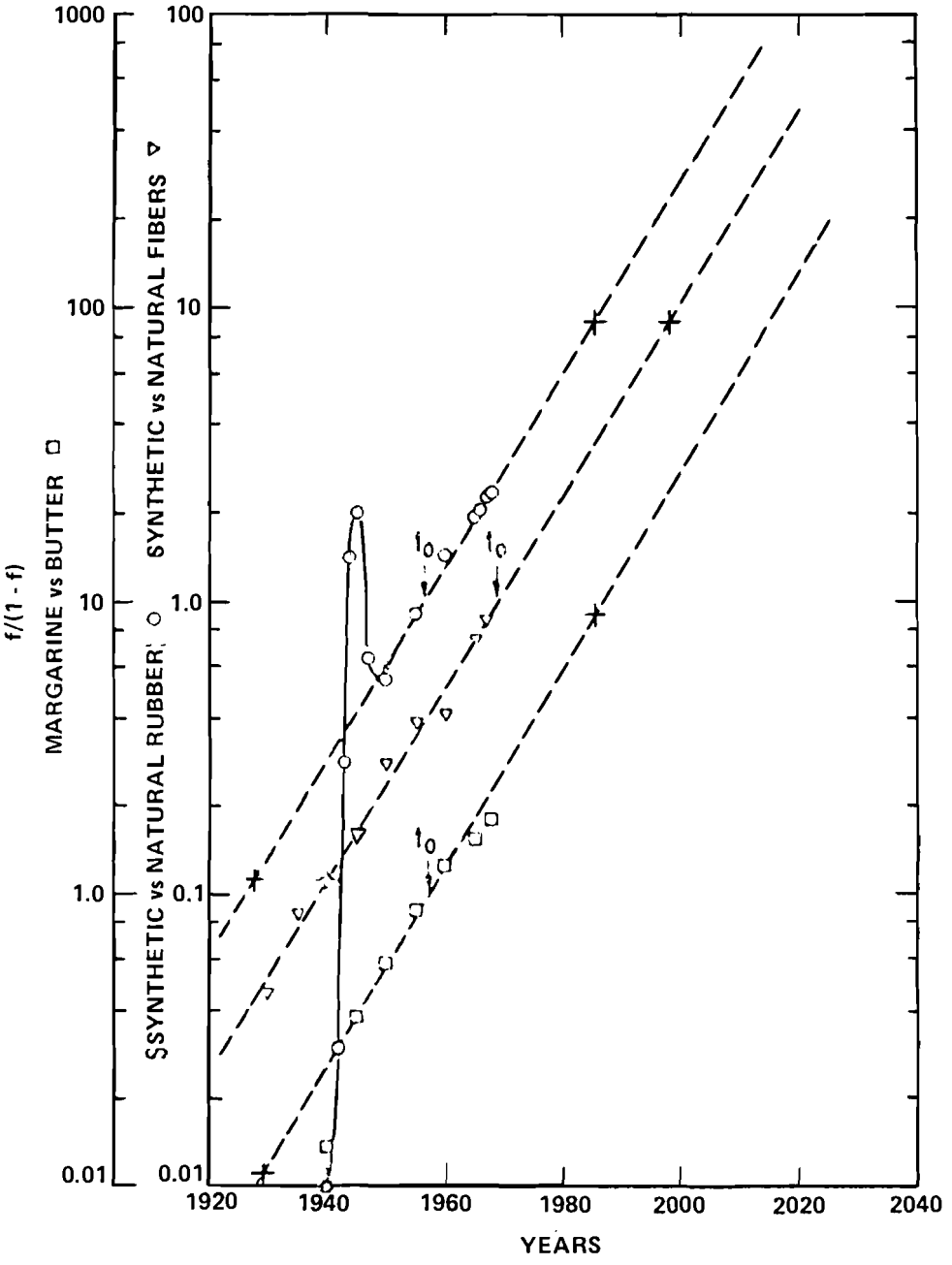
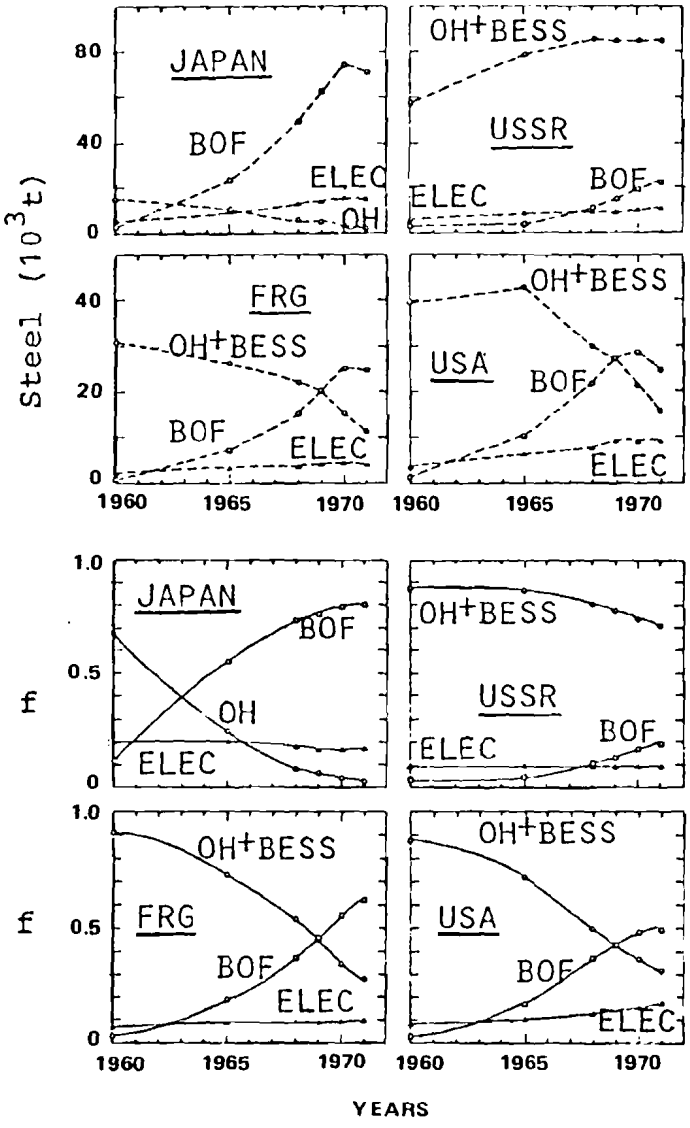


Figure 3. Substitution of three agriculture-based products in the USA.

Source: [3]



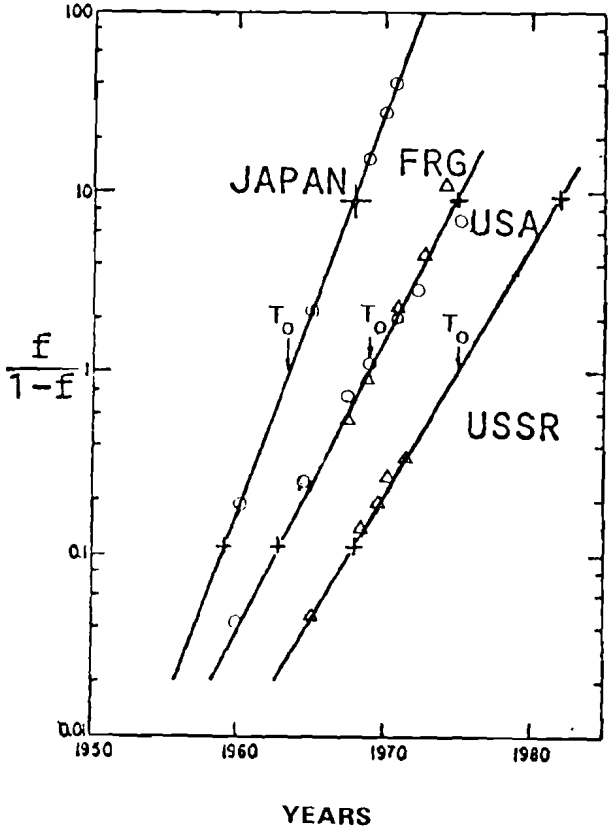


Figure 4. Substitution of BOF for open hearth (OH) and Bessemer steel production in Japan, the USSR, the FRG, and the USA.

Source: [4]

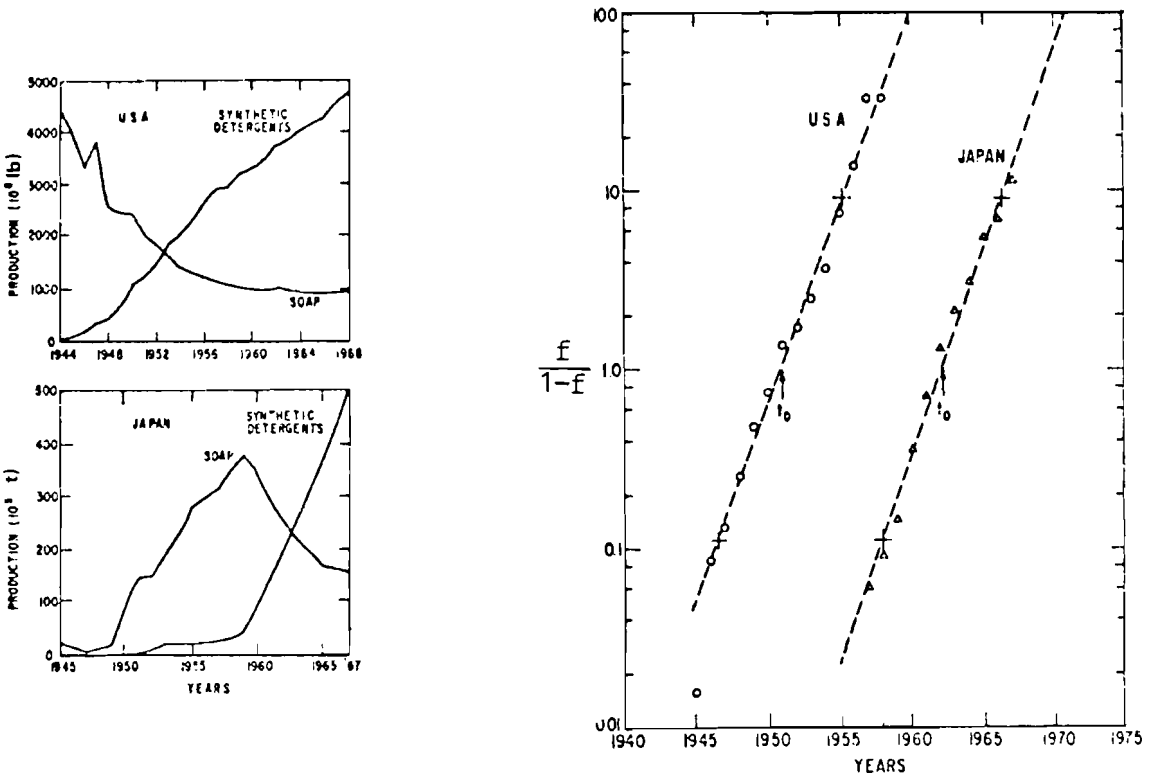


Figure 5. Substitution of synthetic detergents for soap in the USA and Japan.

Source: [4]

Suppose that we start to count the time t at the moment when

$$f_1(0) = f_{10} \quad \text{and} \quad f_2(0) = 1 - f_{10} = f_{20} \ . \tag{2.6}$$

From (2.4) we have

$$\frac{f_{20}}{f_{10}} = e^{-c_{21}t_h} \ ,$$

and the formula (2.2) can be written as follows:

$$f_2(t) = \frac{1}{1 + \frac{f_{10}}{f_{20}} e^{-c_{21}t}} \ . \tag{2.7}$$

The Fisher-Pry model can also be written in the differential form

$$\frac{\dot{f}_2}{f_2} = c_{12}(1 - f_2) \quad . \tag{2.8}$$

The formula (2.7) is actually the solution of this simple non-linear equation of Bernoulli type for the initial conditions (2.6).

The substitution process with two competing commodities may also be described by two differential equations

$$\frac{\dot{f}_1}{f_1} = c_{21}f_2 \quad , \qquad \frac{\dot{f}_2}{f_2} = c_{12}f_1 \quad . \tag{2.9}$$

Notice that the relation

$$\dot{f}_1(t) + \dot{f}_2(t) = 0 \quad ,$$

holds if (2.3) holds. It means that the condition (2.1) is fulfilled for any t if it is fulfilled for one particular t, e.g. t = 0. This observation may seem somewhat redundant in this simple two dimensional case. Its importance will be seen later on when we shall deal with multivariate substitution processes.

Several modifications of the Fisher-Pry model have been suggested [5-10] but they don't seem to be substantial, at least from the point of view of this study.

Marchetti [1,2] brought up the idea of considering different primary energy sources as commodities competing for a market. In the historical period of interest there are at least three or four primary energy sources in the competition and the Fisher-Pry model cannot be directly applied. To handle this case Marchetti, after an analysis of known historical data, suggested a rule called "first in - first out." According to this rule both the technology leaving the market and the newcomer follow the Fisher-Pry straight line (2.5), the former with positive the latter with negative slope, while the fraction of the oldest among the growing ones is determined as a complement to 1. In

this way he was able, using data before 1940, to predict the fractional market share of oil consumption in the USA up to 1970 with a precision better than one percent. See Figure 6.

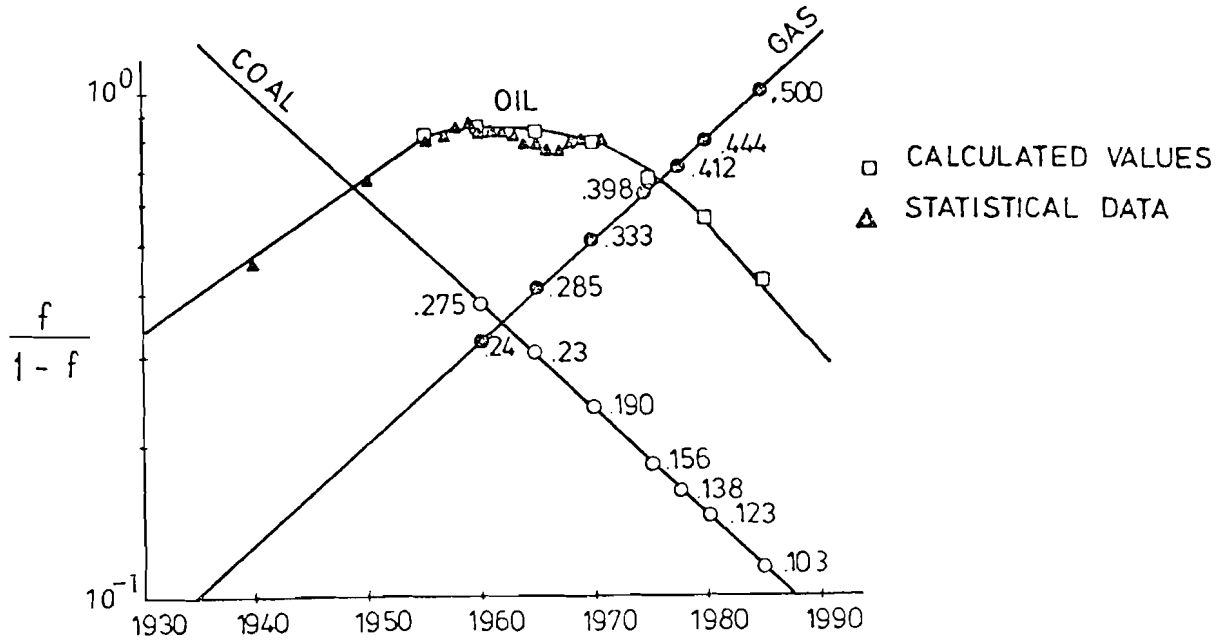


Figure 6. US oil energy fraction calculated from 1930-1940 trend lines.

Source: [2]

When we try to summarize the present knowledge of the substitution processes, mostly based on experience, several questions arise naturally.

(1) The fractions of market share exhibit a much higher regularity than the absolute values of particular productions. Why is it so?

(2) The equation (2.8) says that "the fractional rate of fractional substitution of new for old is proportional to the remaining amount of the old left to be substituted" [3]. This was asserted by Fisher and Pry as a basic assumption. How can such an assumption be justified? Obviously, it cannot be true for more than two competing commodities.

(3) The parameter $c_{21} = -c_{12}$ in equations (2.2) and (2.9) is a characteristic of the difference in quality of two competitors. From the second equation in (2.9) it can be seen that the newcomer f_2 never can penetrate the market if $c_{21} < 0$. Even if $f_{20} > 0$, it will die out. (Remember the competition between airships and airplanes at the beginning of this century.) What does this coefficient depend on?

(4) The previous question was partially answered by Mansfield [9] who showed that the rate constant was positively correlated with profitability of the new technology and negatively influenced by the relative capital investment needed to introduce the new technology. The question whether this relation can be established quantitatively is of extraordinary importance. If the answer were positive it would not be necessary to wait for historical data and the chance of the new technology could be evaluated in advance and also the evolution of the competition could be precalculated given the time instance when the new technology is introduced. For instance, it would be possible to determine under which conditions solar energy may enter the market and what role it will play.

(5) Considering the case of more than two competing technologies it does not seem likely that the evolution of the looser and the newcomer could be entirely independent of the competitor being in transition. Apparently, the Marchetti rule "first in - first out" is a well working approximation of a more general law. What is this law?

(6) The existing substitution models are fatalistic in the sense that they project the future as uniquely predetermined by the past history. An interesting discussion on technological fate can be found in [2]. Is this fate inevitable? A positive answer can hardly be accepted in general. What can be done if the normal competitive technological evolution would lead to drastic ecological changes or if it would threaten the existence of mankind

itself? What is the best way to control the substitution processes?

All these questions will be addressed and hopefully also answered, at least partially, in the following sections.

3. BASIC EQUATION

Let $P_i(t)$ be the production of the ith competing commodity in time t. By production we mean the number of units of the particular commodity produced in the unit of time. What unit is chosen to measure the production is not important at this moment. For instance the steel production can be measured in tons per day, in the case of electricity a megawatt can be chosen as a unit of production.

Consider a finite time interval in which the production was increased from $P_i(t)$ to $P_i(t+\Delta t)$. To realize this increase of production a certain investment was necessary. Let this investment be

$$\alpha_i [P_i(t+\Delta t) - P_i(t)] \quad ,$$

where α_i is the capital needed to increase the production by a unit and will be called specific investment. In α_i also the investment for distribution of the product is respected. Any unit can be chosen to measure the capital and/or investment. The reader may consider a monetary measure if he wants. Later on it will be seen that only ratios are important.

The investment must be covered from some capital sources. One of the possible sources is the capital accumulated by the producer during the time period

$$\int_t^{t+\Delta t} P_i(t)[p(t) - c_i]dt \quad ,$$

where c_i are specific production costs and p(t) is the market

price. By specific production costs we mean all expenses connected with production of a unit of the given commodity including amortization of the material goods used in production and eventually also the tax set on the product by the government.

Let $Q_i(t,\Delta t)$ be the external capital which was extended to the producer from outside. It also may be, for instance, a governmental support given to the producer, if there is some public interest in the production of the given commodity.

Making a balance, we can write

$$\alpha_i[P_i(t+\Delta t) - P_i(t)] = \int_t^{t+\Delta t} P_i(t)[p(t) - c_i]dt + Q_i(t,\Delta t) + \Delta_i(t,\Delta t) \tag{3.1}$$

where $\Delta_i(t,\Delta t)$ is either the part of the capital which was kept by the producer to be invested in the near future, in this case Δ_i is negative, or it is a part of the capital which was accumulated in the past and is invested in the time interval under consideration.

The equation (3.1) can be rewritten into the following form:

$$\int_t^{t+\Delta t} \left\{\alpha_i \dot{P}_i(t) - P_i(t)[p(t) - c_i] - q_i(t)\right\} dt = \Delta_i(t,\Delta t) \quad , \tag{3.2}$$

where

$$\dot{P}_i(t) = \frac{d}{dt} P_i(t)$$

and $q(t)$ is the external capital flow defined by the relation

$$Q_i(t,\Delta t) = \int_t^{t+\Delta t} q(t)dt \quad .$$

When t and/or Δt are changing the right-hand side of equation (3.2) takes different values, positive and negative. As we are interested only in long term behavior of the process studied, $\Delta_i(t,\Delta t)$ in (3.2) can be considered as a random variable with zero mean and instead of (3.2) we can write

$$\int_t^{t+\Delta t} \left\{ \alpha_i \dot{P}_i(t) - P_i(t)[p(t) - c_i] - q_i(t) \right\} dt = 0 \quad . \qquad (3.3)$$

Strictly taken, the variable $P_i(t)$ in (3.2) should be distinguished from the equally denoted variable in (3.3). In (3.2) it means a realization of a random process while in (3.3) it is used to describe an abstract "smoothed" process which is of main interest from the viewpoint of our goals. The stochastic nature of the true process will be considered in more detail later on when we shall deal with estimation of model parameters from historical data and with uncertainty of forecasting.

As the integral in (3.3) is equal to zero for any t and Δt the following differential equation must hold

$$\alpha_i \dot{P}_i(t) = P_i(t)[p(t) - c_i] + q_i(t) \quad . \qquad (3.4)$$

This is the basic equation we shall deal with.

The equation (3.4) contains the market price p(t) which is a very uncertain variable depending on many external and often unknown factors. It should be stressed that we have introduced this quantity only as a separator to be able to formulate the economical balance separately for different competing productions of commodities satisfying the same or similar need. It will be eliminated and will never enter our model. This is perhaps the main trick of the further development of the substitution model.

In the case of societies with planned economies, where no open market exists, the market price p(t) can be understood as a social value of the considered commodity for the present deployment of technologies.

4. MULTIVARIATE COMPETITION

In this section we shall consider the situation when n producers produce commodities satisfying the same need but they use different technologies. We shall derive a multivariate substitution model, a special case of which, for n = 2, is the Fisher-Pry model reported in Section 2.

To start a production with a new technology some investment must be made using external capital sources. No technology can start from zero without external help. This can be clearly seen from equation (3.4). However, if the new technology has been already established and is viable, it must be able to live and grow on its own account. In this section we shall assume that none of the competing technologies is permanently supported by external capital. These conditions can be defined mathematically in the following way.

$$q_i(t) = 0 \quad , \quad \forall i \quad , \quad t > t_0 \quad . \tag{4.1}$$

For n competing technologies we have n equation of type (3.4)

$$\alpha_i \dot{P}_i(t) = P_i(t)[p(t) - c_i]; \quad i = 1,2,\ldots,n \quad . \tag{4.2}$$

Because of the unknown quantity p(t) the system of differential equations (4.2) is not a complete description of the substitution process. We have n equations for n+1 unknowns. Before we show how this difficulty can be overcome two comments are in order.

The differential equation (3.4) has been derived for a growing production. In the system (4.2) all of the competing productions may grow if the demand is growing fast enough. See, for instance, the competition between soap and synthetic detergents in Japan between 1950 and 1957 in Figure 5. However, in general, some productions may subside in the course of competition and the question is whether the corresponding equation in the system (4.2) applies also to this case. The answer is yes under the condition that the production does not fall faster than is the natural amortization of the equipment. Further on we shall assume that

this condition is fulfilled. Then the negative left-hand side of equation (4.2) represents the capital flow which is saved because a part of the worn-out equipment is not renewed. The old technology lives from his stock.

The second comment concerns the market price. In the system of equations (4.2) it was tacitly assumed that the market price is the same for all competing commodities. This is a reasonable assumption if the commodities are in every respect equal. This is, for instance, the case when the same steel is produced by different technologies (e.g. open hearth or Bessemer) or when electricity is generated using different primary energy sources. However, even when competing commodities satisfy the same need, in the sense that the consumer having purchased one commodity will not purchase the other one, they may satisfy this need in qualitatively different ways. The consumer is ready to pay a higher price for a higher quality. Consider, for instance, domestic heating. Both coal and oil can be used to heat a house but oil heating is more comfortable and many of us prefer it even if it is more expensive. Mechanical and electronic wrist watches satisfy the same need, nevertheless many people are ready to pay a higher price for an electronic watch because they want to have it for some psychological reasons. A drastic example of this kind are women's clothes.

To be able to handle at least some of these cases we have to introduce a reference price. As a reference price the price of any of the competing commodities can be chosen. For the sake of simplicity let us choose the lowest price. Let p be this reference price, p_i the price of the ith commodity and Δp_i the difference the consumer is ready to pay for the higher quality. The difference between the price and the specific production costs, which is the source of the capital the producer can accumulate, can be written in the following way

$$p_i - c_i = p + \Delta p_i - c_i = p - (c_i - \Delta p_i) \quad .$$

This shows that the system of differential equations (4.2) holds also for the case of different prices if the specific production

costs are reduced by Δp_i. From now on c_i will mean specific production costs corrected in this way.

To proceed in the development of our model we shall rearrange the equation (4.2) in the following way:

$$\alpha_i \frac{\dot{P}_i}{P_i} = p - c_i$$

$$\alpha_i \frac{d}{dt} (\ell n\, P_i) = p - c_i \quad . \tag{4.3}$$

The same equation can be written for the commodity indexed by j.

$$\alpha_j \frac{d}{dt} (\ell n\, P_j) = p - c_j \quad . \tag{4.4}$$

Subtracting (4.3) from (4.4) we obtain

$$\alpha_j \frac{d}{dt} (\ell n\, P_j) - \alpha_i \frac{d}{dt} (\ell n\, P_i) = c_i - c_j \quad . \tag{4.5}$$

In this way n-1 independent equations can be constructed which do not contain the unknown variable p; however, one equation is still missing.

To begin with a simple case we shall assume that the specific investments are the same for all of the competing technologies:

$$\alpha_i = \alpha \quad \forall i \quad . \tag{4.6}$$

At first sight, it may seem that this is a drastic and very restricting assumption. However, later on when this assumption will be removed, we shall see that the dynamics of market penetration is not very sensitive with respect to this assumption and that the result obtained under this simplification may be a very good approximation for the more general case.

Under the assumption (4.6) the left-hand side of (4.5) can be rearranged in the following way.

$$\alpha_j \frac{d}{dt}(\ell n\, P_j) - \alpha_i \frac{d}{dt}(\ell n\, P_i) = \alpha \frac{d}{dt}\left(\ell n \frac{P_j}{P_i}\right) = \alpha \frac{d}{dt}\left(\ell n \frac{P_j/P}{P_i/P}\right) \tag{4.7}$$

$$= \alpha \frac{d}{dt}\left(\ell n \frac{f_j}{f_i}\right) \quad , \tag{4.8}$$

where P is the total production of competing commodities

$$P = \sum_{i=1}^{n} P_i \tag{4.9}$$

and f_i is the fractional market share

$$f_i = \frac{P_i}{P} \quad . \tag{4.10}$$

Using (4.8) the equation (4.5) can be rewritten for market shares.

$$\frac{d}{dt}\left(\ell n \frac{f_j}{f_i}\right) = \frac{c_i - c_j}{\alpha} \tag{4.11}$$

In this way the market price p(t) has been eliminated. The equation (4.11) can be written for any i and j; however, only n-1 of these equations are independent. Hence we still have a system of n-1 equations for n unknowns f_i, i = 1,2,...,n. But having passed from absolute values of productions P_i to corresponding market shares f_i we have the possibility to make use of an additional equation, which is independent of (4.11):

$$\sum_{i=1}^{n} f_i(t) = 1 \quad . \tag{4.12}$$

Now we have a complete system of differential equations the solution of which, for given initial conditions, is unique.

The system of equations (4.11) and (4.12) is somewhat unsymmetric. It would be good to have it in a symmetric form. Let us find such a form.

The equation (4.11) can be reorganized in the following way.

$$\frac{\dot{f}_j}{f_j} + \frac{c_j}{\alpha} = \frac{\dot{f}_i}{f_i} + \frac{c_i}{\alpha} \quad .$$

As this relation holds for any j and i there must exist a function $\varphi(t)$ which is common for all components of the system and for which

$$\frac{\dot{f}_j}{f_j} + \frac{c_j}{\alpha} = \varphi(t) \quad , \qquad \forall j \quad . \tag{4.13}$$

Let us determine this function. From (4.12)

$$\sum_{j=1}^{n} \dot{f}_j(t) = 0 \quad . \tag{4.14}$$

From (4.13) we have

$$\dot{f}_j = f_j\left[\varphi(t) - \frac{c_j}{\alpha}\right] \quad ,$$

and after the substitution in (4.14) we get

$$\varphi(t) = \frac{1}{\alpha} \sum_{j=1}^{n} c_j f_j \quad . \tag{4.15}$$

Now, the symmetric system can be obtained from (4.13):

$$\dot{f}_i = \frac{1}{\alpha} f_i\left(\sum_{j=1}^{n} c_j f_j - c_i\right) \quad , \qquad \forall i \quad . \tag{4.16}$$

Making use of the obvious relation

$$c_i = \sum_{j=1}^{n} c_i f_j$$

and introducing the notation

$$c_{ij} = \frac{c_i - c_j}{\alpha} \tag{4.17}$$

$$c_{ii} = 0 \quad ,$$

the equation (4.16) can be written as

$$\dot{f}_i + f_i \sum_{j=1}^{n} c_{ij} f_j = 0 \quad , \quad \forall i \quad . \tag{4.18}$$

Notice that this system of equations also holds when the coefficients c_{ij} are time-dependent. No assumptions have been made in this respect. Notice also that for n=2 the system (4.18) is

$$\dot{f}_1 + c_{12} f_1 f_2 = 0 \quad , \qquad \dot{f}_2 + c_{21} f_1 f_2 = 0 \quad ,$$

which is the Fisher-Pry model in the differential form (2.9).

Only very rarely an explicit solution of a system of non-linear differential equations can be found. Fortunately, system (4.18) is an exception. A general solution can be found in different ways. We shall proceed in a way which is somewhat tricky but simple.

Evidently for any t

$$f_i(t) = \frac{f_i(t)}{\sum_{j=1}^{n} f_j(t)}$$

$$f_i(t) = \frac{1}{1 + \sum_{j \neq i} \frac{f_j(t)}{f_i(t)}} \quad . \tag{4.19}$$

With notation (4.17) the equation (4.11) reads

$$\frac{d}{dt}\left(\ell n \frac{f_j(t)}{f_i(t)}\right) = c_{ij} = -c_{ji}$$

and can be easily integrated:

$$\ell n \frac{f_j(t)}{f_i(t)} - \ell n \frac{f_{jO}}{f_{iO}} = -\int_{t_O}^{t} c_{ji}(\tau)d\tau \qquad (4.20)$$

$$\frac{f_j(t)}{f_i(t)} = \frac{f_{jO}}{f_{iO}} \exp\left\{-\int_{t_O}^{t} c_{ji}(\tau)d\tau\right\} \quad , \qquad (4.21)$$

where

$$f_{iO} = f_i(t_O) \quad , \qquad \forall\, i$$

are the initial conditions. It only remains to substitute (4.21) into (4.19).

$$f_i(t) = \frac{1}{1 + \sum\limits_{j\neq i} \frac{f_{jO}}{f_{iO}} \exp\left\{-\int_{t_O}^{t} c_{ji}(\tau)d\tau\right\}} \quad . \qquad (4.22)$$

If it can be assumed that the coefficients c_{ij} defined by (4.17) are time invariant the following simple result is obtained:

$$f_i(t) = \frac{1}{1 + \sum\limits_{j\neq i} \frac{f_{jO}}{f_{iO}} e^{-c_{ji}(t - t_O)}} \quad . \qquad (4.23)$$

If we denote

$$k_{ji} = \ell n \frac{f_{jO}}{f_{iO}} \quad , \qquad (4.24)$$

formula (4.23) becomes

$$f_i(t) = \frac{1}{1 + \sum_{j \neq i} e^{k_{ji} - c_{ji}(t - t_o)}} \quad . \tag{4.25}$$

Notice that for n=2, formula (4.23) gives the Fisher-Pry model in the form (2.7) while (4.25) corresponds to (2.2).

The simple formula (4.23) assumes that the coefficients c_{ji} are time invariant. According to (4.17) these coefficients are determined by ratios of specific production costs (eventually corrected by Δp_i as discussed above) and specific investment. It is known [11] that due to the learning effect the efficiency of the direct labor input improves with the number of units produced. This should be reflected in specific production costs and also in specific investments. Therefore the assumption that their ratio remains constant does not seem unrealistic at least from the time when the technology was well established. Moreover, from the way that the time-varying coefficients c_{ji} enter the more general formula (4.22) it can be seen that a mean value, defined as

$$c_{ji} = \frac{1}{T} \int_{t_o}^{t_o+T} c_{ji}(\tau) d\tau \quad ,$$

may serve as a good approximation for $t_o \leq t \leq t_o+T$ if $c_{ij}(\tau)$ does not vary too drastically.

Example 1. Substitution of primary energy sources

Now it will be shown how the multivariate substitution model, derived in this section, works in a practical example. The model will be applied to describe the competition between different primary energy sources during the past 110 years. Wood, coal, oil and natural gas are considered as competitors in the world consumption of energy.

Under the assumption that c_{ij} are constants for all j and i the relation (4.20) can be written in the following form

$$\ln \frac{f_j(t)}{f_i(t)} = k_{ji} - c_{ji}(t - t_o) \quad . \tag{4.26}$$

It indicates that the logarithms of the ratios of market shares for all pairs of competitors, when plotted as functions of time, should follow straight lines. Figures 7a, b, c, d show that in the given example it is true with a very good approximation. The straight lines in these figures were obtained as least squares fits. Notice that oil and natural gas are equivalent competitors in a certain sense. This can be seen from horizontal lines in Figures 7c, d and from parallelism of lines for oil and natural gas in Figures 7a, b.

To see how the model can be used for forecasting, only the historical data between 1930 and 1950 were taken to estimate the model parameters. The projections obtained in this way for the future (and also for the past) are given in Figure 8 where also true historical data are registered for comparison. The figure also shows how the future, not yet known, development of the natural competition could be if no new and superior technology -- like nuclear -- were introduced. The historical data given in this example were collected at IIASA by N. Nakicenovic [12] who also performed this preliminary calculation. The question of estimation of model parameters from historical data will be discussed in Section 6 where a more detailed analysis and description of this example will be given. The problem of forecasting is studied in Section 7.

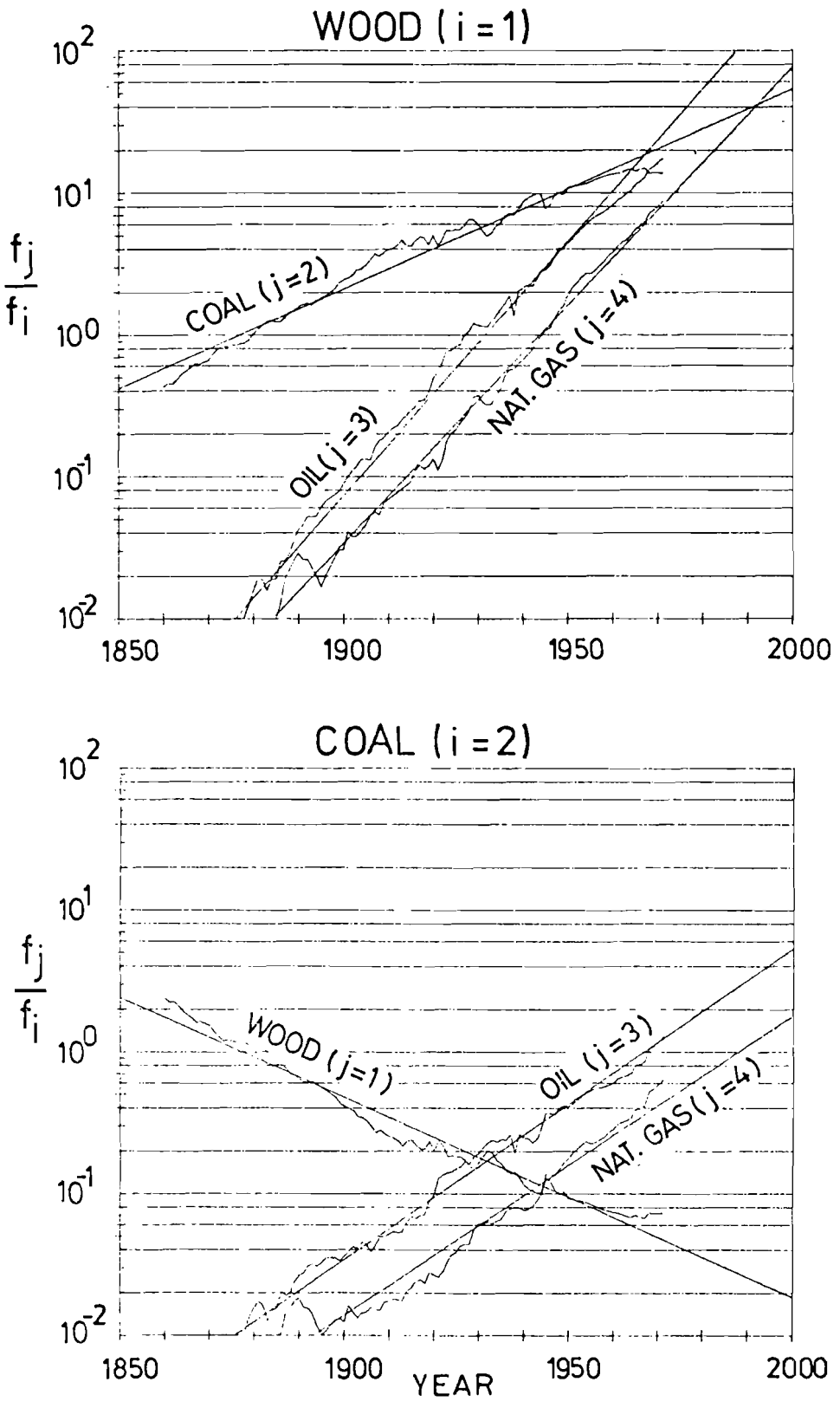


Figure 7a,b. Ratios of market shares plotted vs. time in semilogarithmic scale.

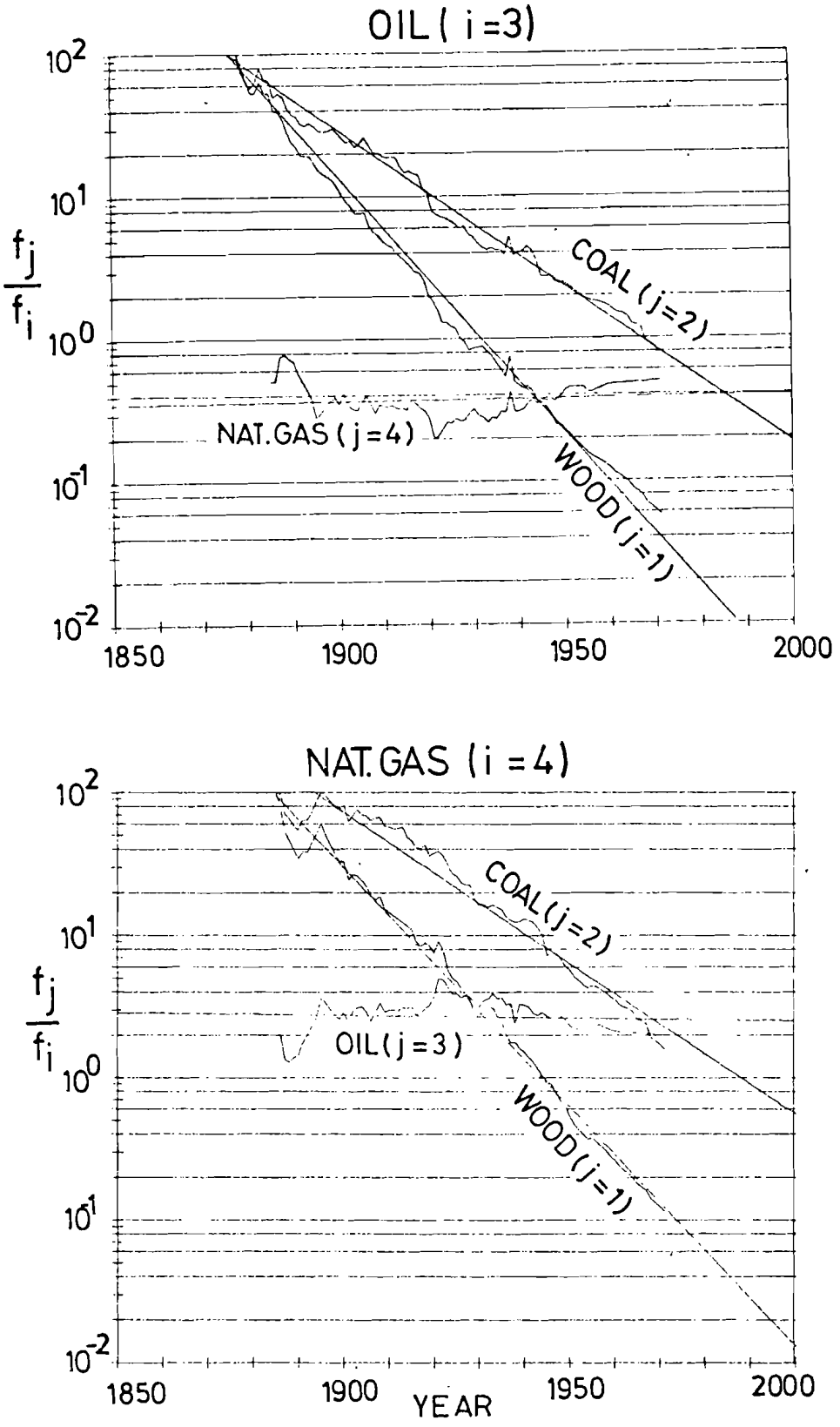


Figure 7c,d. (continued)

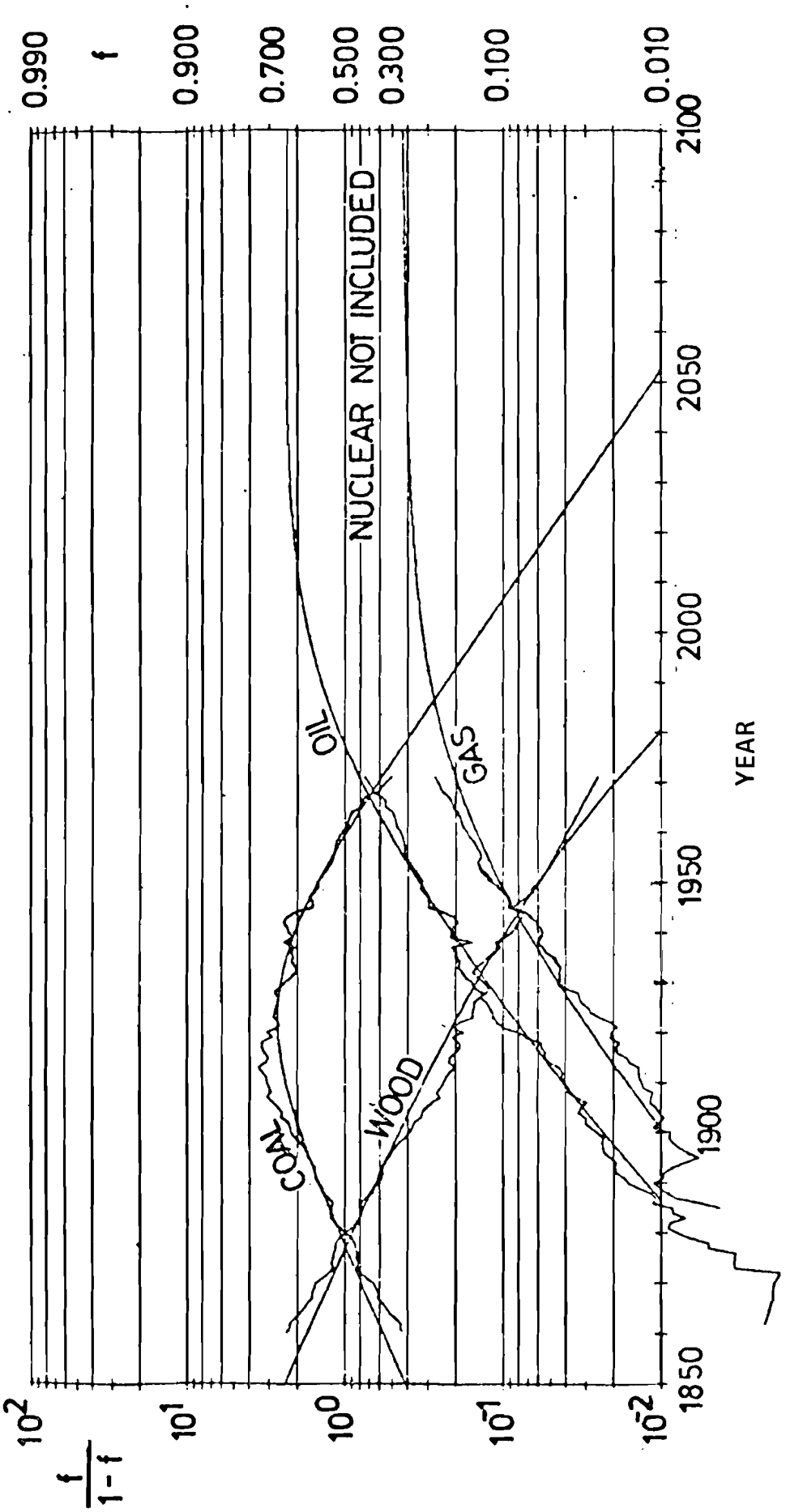


Figure 8. World energy data and fit to substitution model. Only data from time period 1930-1950 used for parameter estimation.

5. COMPETITION UNDER DIFFERENT SPECIFIC INVESTMENTS

In the previous section a multivariate substitution model has been developed under the assumption that the specific investments α_i, i = 1,2,...,n, were the same for all of competing technologies. This simplifying assumption will now be removed and it will be shown that the overall dynamics of the substitution does not change much when the specific investments of competing technologies are slightly different. As a matter of fact, this was already indicated by the example of primary energy substitution where the specific investments apparently are not equal, and also by many examples of two-dimensional competition given by Fisher and Pry [3,4]. Now we shall investigate this favorable feature of the model in detail.

We shall start our investigation with the equation (4.5) which can be written in the following form

$$\alpha_j \frac{\dot{P}_j}{P_j} + c_j = \alpha_i \frac{\dot{P}_i}{P_i} + c_i \quad . \tag{5.1}$$

For n competing technologies n-1 independent equations of this type can be written. Introducing the fractional market shares (4.10) we have

$$\frac{\dot{P}_j}{P_j} = \frac{\dot{f}_j}{f_j} + \frac{\dot{P}}{P} \quad , \tag{5.2}$$

where P is the total production of all competing technologies (4.9). If the growth rate factor

$$\rho = \frac{\dot{P}}{P} = \frac{d}{dt}(\ln P) \tag{5.3}$$

is introduced the equation (5.1) gets the form

$$\alpha_j \frac{\dot{f}_j}{f_j} + c_j + \alpha_j \rho = \alpha_i \frac{\dot{f}_i}{f_i} + c_i + \alpha_i \rho \quad . \tag{5.4}$$

By the growth rate ρ (5.3) a new exogenous parameter, or possibly a variable, is introduced. The influence of this exogenous quantity on the system dynamics will be discussed in more detail later

on; at this moment notice only that the growth rate ρ can be cancelled in (5.4) if $\alpha_j = \alpha_i$. Notice also that all parameters in equation (5.4) can be time-varying in general. Some assumptions concerning this point will be made in due course.

As the relation (5.4) holds for any pair of indices i and j there exists a function $\varphi(t)$ which is common for all of the competitors and for which

$$\alpha_i \frac{\dot{f}_i}{f_i} + c_i + \alpha_i \rho = \varphi \quad ; \quad \forall i \quad . \tag{5.5}$$

Similarly to Section 4 the function $\varphi(t)$ can be determined from the condition

$$\sum_{i=1}^{n} f_i = 1 \tag{5.6}$$

$$\sum_{i=1}^{n} \dot{f}_i = 0 \quad . \tag{5.7}$$

From (5.5)

$$\dot{f}_i = f_i\left(\frac{1}{\alpha_i}\varphi - \frac{c_i}{\alpha_i} - \rho\right) \quad , \tag{5.8}$$

and using (5.7) the following expression for $\varphi(t)$ is obtained:

$$\varphi = \frac{\sum_{j=1}^{n} f_j\left(\frac{c_j}{\alpha_j} + \rho\right)}{\sum_{j=1}^{n} \frac{f_j}{\alpha_j}} \quad . \tag{5.9}$$

The replacement of φ in (5.8) by (5.9) and a simple rearrangement gives the following symmetrical system of differential equations:

$$\dot{f}_i + f_i \frac{\sum_{j=1}^{n} c_{ij} f_j}{\sum_{j=1}^{n} a_{ij} f_j} = 0 \quad ; \quad i = 1,2,\ldots,n \quad , \tag{5.10}$$

where

$$a_{ij} = \frac{\alpha_i}{\alpha_j} \tag{5.11}$$

$$c_{ij} = \frac{\alpha_i}{\alpha_j}\left(\frac{c_i}{\alpha_i} + \rho\right) - \left(\frac{c_j}{\alpha_j} + \rho\right) = \frac{c_i - c_j}{\alpha_j} + \left(\frac{\alpha_i}{\alpha_j} - 1\right)\rho \,. \tag{5.12}$$

Notice that the following relations hold for the parameters c_{ij} and a_{ij} entering the system of differential equations (5.10) by which the multivariate substitution process is governed.

$$a_{ij} = \frac{1}{a_{ji}} = \frac{a_{ir}}{a_{jr}} = \frac{a_{rj}}{a_{ri}} \tag{5.13}$$

$$a_{ii} = 1 \tag{5.14}$$

$$c_{ij} = -a_{ij}\, c_{ji} = c_{rj} - a_{ij}\, c_{ri} = a_{rj}(c_{ir} - c_{jr}) \tag{5.15}$$

$$c_{ii} = 0 \quad . \tag{5.16}$$

This means that the system dynamics is fully determined only by $2(n-1)$ independent parameters (possibly time-varying), for instance, by

$$a_{ri}\,,\, c_{ri} \quad ; \quad \forall\, i \neq r$$

or by

$$a_{ir}\,,\, c_{ir} \quad ; \quad \forall\, i \neq r \quad ,$$

where r is the index of an arbitrarily chosen reference competitor. All remaining parameters are determined by relations (5.13) to (5.16).

Notice also that the system can be considered as a system with constant parameters if $\left(\frac{c_i}{\alpha_i} + \rho\right)$ and the ratios $\frac{\alpha_i}{\alpha_j}$ are time-invariant, i.e. not necessarily c_i and α_i separately.

A comment is in order concerning the growth rate factor ρ. The introduction of this factor is the price we have to pay in order to get rid of the market price p in the case of different specific investments α_i. Actually the growth rate factor ρ is related to the market price p and it is true that eliminating the uncertain quantity p we introduced the other one ρ. However, it

seems to be more advantageous to operate with ρ instead of p for several reasons: (1) From the expression (5.12) it can be seen that the total growth rate factor ρ does not have much influence if the ratio α_i/α_j is close to one. (2) Often the growth of the total production P exhibits an exponential behavior and the factor ρ can be considered as a constant with reasonable approximation. World energy consumption and U.S. total energy consumption are given in Figures 9 and 10 [2] as examples of this kind. In the former case $\rho \approx 0.02$, while for the USA $\rho \approx 0.03$ if one year is taken as the time unit. Later on it will be seen that the dynamics of the system is rather insensitive with respect to stochastic fluctuations of $\rho(t)$ around some mean value even when the specific investments are considerably different. (3) If one would prefer to go deeper into the market relations it would be necessary to introduce and parametrize the market demand function and to make some additional assumption like existence of market equilibrium, etc.

The solution of the system of differential equations (5.10), by which the substitution process is governed in the case of different specific investments α_i, i.e. $a_{ji} \neq 1$, cannot be given in a closed explicit form. A straightforward way to obtain the solution for a particular case is the stepwise numerical solution using several known general algorithms. However, if it can be assumed that the ratios $a_{ij} = \alpha_i/\alpha_j$ are time-invariant for all i and j it is possible to reduce the solution of the system of n nonlinear differential equations (5.10) to a problem of finding the root of a simple univariate transcendental equation. The method we are going to develop, makes it possible to determine the market shares $f_i(t)$, $i = 1,2,\ldots,n$ for any given time instant t in a simple way and it also gives a better picture about the sensitivity of the solution with respect to parameter values and their possible variations.

Choose one of the competing technologies, say with index r, as the reference competitor, divide the relation (5.4) by $\alpha_r > 0$ and subtract from both sides of (5.4) the term $\left(\frac{c_r}{\alpha_r} - \rho\right)$. The following relation equivalent to (5.4) is obtained:

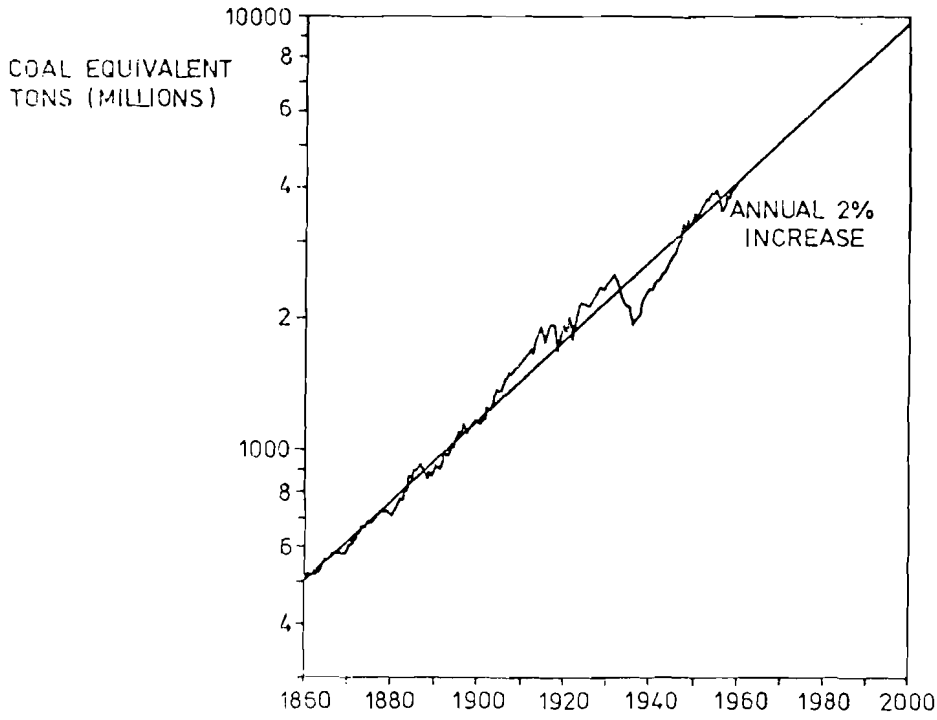


Figure 9. World energy consumption, including wood and farm waste.

Source: [2]

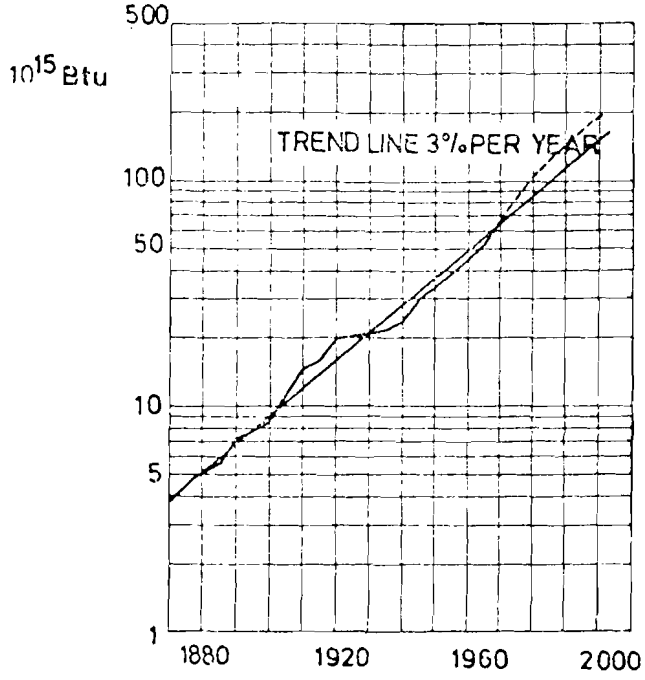


Figure 10. US total energy consumption.

Source: [2]

$$a_{jr}\frac{\dot{f}_j}{f_j} + c_{jr} = a_{ir}\frac{\dot{f}_i}{f_i} + c_{ir} \quad . \tag{5.17}$$

Similarly to (5.4) this relation also holds for any pair of indices i and j and therefore

$$a_{ir}\frac{\dot{f}_i}{f_i} + c_{ir} = \varphi_r \quad , \tag{5.18}$$

where φ_r is a function of time which is common for all indices i. The meaning of this function becomes clear when the index i is chosen as i = r and the relations (5.14) and (5.16) are considered. Then from (5.18)

$$\varphi_r = \frac{\dot{f}_r}{f_r} = \frac{d}{dt}\,(\ell n\, f_r) \quad . \tag{5.19}$$

Dividing the whole equation (5.18) by a_{ir} and using the first equalities in (5.13) and (5.15), we obtain

$$\frac{1}{a_{ir}} = a_{ri} \quad , \qquad \frac{c_{ir}}{a_{ir}} = -\,c_{ri} \quad .$$

The relation (5.18) can be rearranged into the following form

$$\frac{\dot{f}_i}{f_i} = c_{ri} + a_{ri}\varphi_r \quad . \tag{5.20}$$

Integration of (5.20), under the assumption that a_{ri} is time-invariant, gives

$$\ell n\,\frac{f_i(t)}{f_i(t_0)} = \int_{t_0}^{t} c_{ri}(\tau)\,d\tau + a_{ri}\int_{t_0}^{t} \varphi_r(\tau)\,d\tau \tag{5.21}$$

$$f_i(t) = \kappa_{ri}(t)\,e^{a_{ri}\psi_r(t)} \quad , \tag{5.22}$$

where

$$\kappa_{ri}(t) = f_i(t_0)\, e^{\int_{t_0}^{t} c_{ri}(\tau)\,d\tau} \tag{5.23}$$

$$\psi_r(t) = \int_{t_0}^{t} \varphi_r(\tau)\,d\tau = \ln \frac{f_r(t)}{f_r(t_0)} \quad . \tag{5.24}$$

If it can be assumed that the parameters c_{ri} are also time-invariant then the formula (5.23) gets the form

$$\kappa_{ri}(t) = f_i(t_0)\, e^{c_{ri}(t - t_0)} \quad . \tag{5.25}$$

Notice that the variations of the parameters $c_{ri}(t)$ are smoothed by the integral in (5.23). This also shows that the stochastic fluctuations of the growth rate factor ρ, entering c_{ri} according to (5.12), may well be neglected even when α_i, $i = 1,2,\ldots,n$ are considerably different.

For given initial condition $f_i(t_0)$ and given c_{ri} the factor $\kappa_{ri}(t)$ can be easily calculated for all i and any time instance t using (5.23) or (5.25). To be able to determine also the market shares $f_i(t)$ according to (5.22) it is necessary to know the value of the single function $\psi_r(t)$ for the particular t. This value can be determined from the condition

$$\sum_{i=1}^{n} f_i = 1 \quad ,$$

when f_i are considered functions of an unknown value ψ_r for given t. If we introduce the function

$$\xi_r(\psi_r) = \sum_{i=1}^{n} f_i(\psi_r) - 1$$

$$\xi_r(\psi_r) = \sum_{i=1}^{n} \kappa_{ri}\, e^{\frac{\alpha_r}{\alpha_i}\psi_r} - 1 \quad , \tag{5.26}$$

then the unknown value $\psi_r(t)$ is the root of the transdendental equation

$$\xi_r(\psi_r) = 0 \quad . \tag{5.27}$$

As both κ_{ri} and a_{ri} for all i are positive the derivative $d\xi_r/d\psi_r$ is also positive for any ψ_r -- the function $\xi_r(\psi_r)$ is monotonous -- and consequently the real root of the equation (5.27) is unique. It can be found by several well known iterative numerical methods. Before we go into these details some rearrangement of the equation (5.27) is necessary.

The advantage of the procedure outlined above is that it operates with the minimum number of parameters. The disadvantage is that it is unsymmetrical in the sense that it depends on the choice of the reference competitor indexed by r. A more detailed analysis shows that an unsuitable choice of the reference competitor might lead to numerical difficulties. To avoid these possible difficulties we shall forego the minimum number of parameters and we shall modify the procedure to maintain the symmetry.

Let $\bar{\alpha}$ be some mean value of all α's the suitable choice of which will be made later on. Dividing equation (5.5) by $\alpha_i > 0$, we can write it as

$$\frac{d}{dt} \ln f_i = \left(\frac{\varphi}{\bar{\alpha}} - \frac{c_i + \alpha_i \rho}{\bar{\alpha}} \right) \frac{\bar{\alpha}}{\alpha_i} \quad . \tag{5.28}$$

Integration of this equation over the time interval (t_0,t) under the assumption that the ratio $\bar{\alpha}/\alpha_i$ is time-invariant gives

$$\ln \frac{f_i(t)}{f_i(t_0)} = \left(\int_{t_0}^{t} \frac{\varphi}{\bar{\alpha}} dt - \int_{t_0}^{t} \frac{c_i + \alpha_i \rho}{\bar{\alpha}} dt \right) \frac{\bar{\alpha}}{\alpha_i}$$

$$f_i(t) = f_{i_0} e^{(\psi(t) - \beta_i(t)) a_i} \quad , \tag{5.29}$$

where

$$a_i = \frac{\bar{\alpha}}{\alpha_i} \quad , \tag{5.30}$$

$f_{i_0} = f_i(t_0)$, $\psi(t)$ is a function of time which is common for all i but unknown for $t \neq t_0$ and

$$\beta_i(t) = \int_{t_0}^{t} \frac{c_i + \alpha_i \rho}{\bar{\alpha}} \, dt \tag{5.31}$$

or, when $(c_i + \alpha_i \rho)/\bar{\alpha}$ is constant,

$$\beta_i(t) = \frac{c_i + \alpha_i \rho}{\bar{\alpha}} (t - t_0) \quad . \tag{5.32}$$

For any given $t \neq t_0$ all market shares (5.29) can be considered as functions of a single quantity $\psi = \psi(t)$. The correct value of this quantity can be determined as the real root of the equation

$$\xi(\psi) = 0 \quad , \tag{5.33}$$

where

$$\xi(\psi) = \sum_{i=1}^{n} f_i(\psi) - 1$$

$$\xi(\psi) = \sum_{i=1}^{n} f_{i_0} e^{(\psi - \beta_i) a_i} - 1 \quad . \tag{5.34}$$

As all $a_i = \bar{\alpha}/\alpha_i$ can be only positive the derivative

$$\xi'(\psi) = \frac{d\xi(\psi)}{d\psi} = \sum_{i=1}^{n} a_i f_{i_0} e^{(\psi - \beta_i) a_i} \tag{5.35}$$

$$\xi'(\psi) = \sum_{i=1}^{n} a_i f_i(\psi) \quad , \tag{5.36}$$

is also positive and the function $\xi(\psi)$ is monotonously increasing. Therefore the real root of the equation (5.32), we are lokking for, is unique and can be easily found by the Newton-Raphson method illustrated in Figure 11 and realized by the recursive formula

$$\psi^{(k+1)} = \psi^{(k)} - \frac{\xi(\psi^{(k)})}{\xi'(\psi^{(k)})} \quad , \tag{5.37}$$

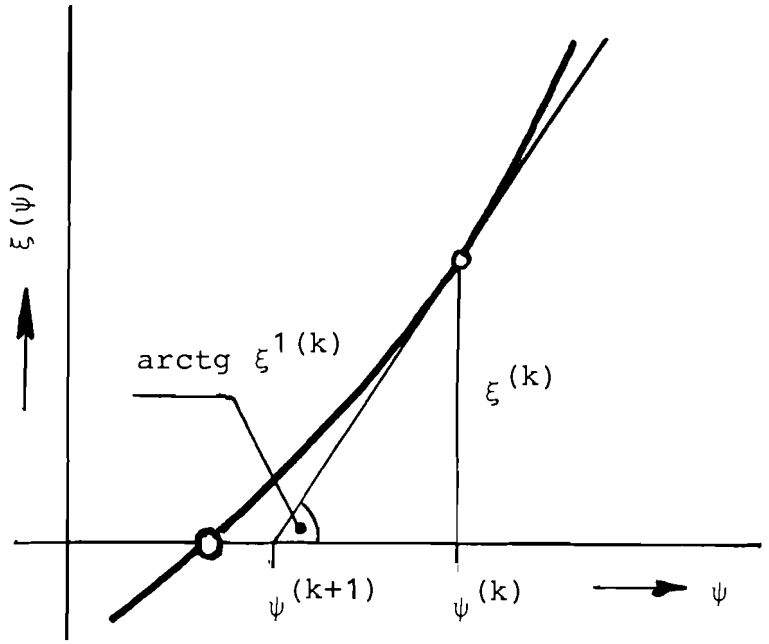


Figure 11. Newton-Raphson method.

where $\psi^{(k)}$ means the kth approximation. Theoretically the root is found as the limit

$$\psi = \lim_{k\to\infty} \psi^{(k)} \quad ;$$

however, in practice only a few iterations are fully sufficient to obtain the root with the required precision if the starting point $\psi^{(o)}$ is well chosen. Hence, the question of suitable choice of the initial approximation $\psi^{(o)}$, which we are going to answer, is of great practical importance. In this context we shall also find a suitable mean value $\bar{\alpha}$ of all α's which has not been defined yet.

In the case of equal α's, i.e. for $a_i = \bar{\alpha}/\alpha_i = 1$, $\forall i$, the root of equation (5.33), i.e. the zero point of the function (5.34), can be calculated explicitly.

In this case, equation (5.33) reads

$$e^{\psi} \sum_{i=1}^{n} f_{i_0} e^{-\beta_i} - 1 = 0$$

and its solution is

$$e^{\psi} = \frac{1}{\sum_{i=1}^{n} f_{i_0} e^{-\beta_i}} \tag{5.38}$$

$$\psi = - \ell n \left(\sum_{i=1}^{n} f_{i_0} e^{-\beta_i} \right) \quad . \tag{5.39}$$

Substitution of $a_i = 1$ and (5.38) into formula (5.29) gives

$$f_i(t) = \frac{f_{i_0} e^{-\beta_i(t)}}{\sum_{j=1}^{n} f_{j_0} e^{-\beta_j(t)}} = \frac{1}{1 + \sum_{j \neq i} \frac{f_{j_0}}{f_{i_0}} e^{\beta_i(t) - \beta_j(t)}} , \tag{5.40}$$

which is the solution we obtained for this simple case in Section 4.

The value of ψ given by formula (5.39) can well serve as the initial approximation of the root of equation (5.33) if the mean value $\bar{\alpha}$ is chosen in such a way that the ratios $\bar{\alpha}/\alpha_i = a_i$ are as close to one as possible. To meet this requirement we choose $\bar{\alpha}$ so that it minimizes the expression

$$\sum_{i=1}^{n} \left(\frac{\bar{\alpha}}{\alpha_i} - 1 \right)^2 \quad .$$

The mean value which has this property is

$$\bar{\alpha} = \frac{\sum_{i=1}^{n} \frac{1}{\alpha_i}}{\sum_{i=1}^{n} \frac{1}{\alpha_i^2}} \quad . \tag{5.41}$$

The choice of the initial approximation $\psi^{(o)}$ according to (5.39) and of the mean value $\bar{\alpha}$ according to (5.41) is not necessarily the best one for some particular case. Nevertheless, inspite of its simplicity it appeared to be fully convenient.

To complete the numerical algorithm one last question remains to be answered: How is it possible to recognize that the required precision has been reached and the iteration (5.37) can be stopped?

Let ψ be the true root of equation (5.33), $\psi^{(k)}$ its kth approximation, $f_i^{(k)}(t)$ the kth approximation of $f_i(t)$ and

$$\delta\psi^{(k)} = \psi^{(k)} - \psi$$

$$\delta f_i^{(k)} = f_i^{(k)} - f_i$$

the corresponding errors. From (5.29)

$$f_i^{(k)} = f_{i_0} e^{(\psi^{(k)} - \beta_i)a_i} = f_{i_0} e^{(\psi - \beta_i)a_i} e^{\delta\psi^{(k)} a_i} = f_i e^{\delta\psi^{(k)} a_i}$$

$$\delta f_i^{(k)} = f_i \left(e^{\delta\psi^{(k)} a_i} - 1 \right) \quad .$$

As all a_i are positive all errors $\delta f_i^{(k)}$ have the same sign. From (5.34) we have

$$\sum_{i=1}^{n} \delta f_i^{(k)} = \xi(\psi^{(k)})$$

and consequently

$$\left| \max_i \delta f_i^{(k)} \right| \leq |\xi(\psi^{(k)})| \quad .$$

This means that if ε is the greatest acceptable error in the calculation of $f_i(t)$ and the iteration (5.37) is stopped when $|\xi| \leq \varepsilon$, then the required precision is guaranteed for all i.

The numerical solution can be summarized into the following.

A l g o r i t h m PENETR:

1. For given α_i, i = 1,2,...,n, calculate the mean $\bar{\alpha}$ according to (5.41) and a_i for all i according to (5.30).

2. For given c_i, ρ and $(t - t_0)$ calculate β_i for all i according to (5.32).

3. For given $f_i(t_0) = f_{i0}$ calculate the starting value $\Psi^{(k)}$, k = 0, according to (5.39).

4. Assuming $\Psi = \Psi^{(k)}$ calculate the market shares $f_i = f_i(t)$ for all i according to (5.29).

5. Calculate $\xi = \sum_{i=1}^{n} f_i - 1$. If the absolute value $|\xi|$ is less than the maximum acceptable error in the calculation of f_i stop the calculation.

6. Calculate new $\Psi^{(k+1)}$ using (5.37) together with (5.36) and repeat 4.

This algorithm is realized by the FORTRAN subroutine PENETR (N,C,AL,FØ,T,F) the listing of which can be found in Appendix C.

Algorithm PENETR operates with 2n + 1 parameters, namely c_i, α_i, i = 1,2,...,n, and ρ. However, as was shown above, the solution $f_i(t)$ for given $f_i(t_0)$ is uniquely determined only by 2(n-1) parameters. Therefore 3 parameters of the algorithm PENETR are redundant and could be removed. It is easy to verify that the same result is obtained if the original parameters ρ, α, c are substituted by the modified parameters $\tilde{\rho}$, $\tilde{\alpha}$, $\tilde{c}$ determined in the following way:

$$\begin{aligned} \tilde{\rho} &= 0 \\ \tilde{\alpha}_r &= 1 \\ \tilde{\alpha}_i &= a_{ir} = \frac{1}{a_{ri}} = \frac{\alpha_i}{\alpha_r} \end{aligned} \qquad (5.42)$$

$$\tilde{c}_r = 0$$

$$\tilde{c}_i = c_{ir} = -\frac{c_{ri}}{a_{ri}} = \frac{c_i - c_r}{\alpha_r} + \left(\frac{\alpha_i}{\alpha_r} - 1\right)\rho \quad , \tag{5.43}$$

where r is the index of an arbitrarily chosen reference competitor. Of course, other modifications of parameters are also possible.

To demonstrate the application of the algorithm PENETR we shall give a simple example. Further examples will be given in the next sections where we shall deal with real practical cases.

Example 2. Sensitivity analysis of two-dimensional competition with respect to different specific investments.

Consider a two-dimensional competition where f_2 is the market share of the new, winning technology. The evolution of the market shares is described by the differential equation (5.10) which for n = 2 and i = 2 reads

$$\dot{f}_2 + f_2 \frac{c_{21} f_1}{a_{21} f_1 + f_2} = 0 \quad ,$$

or, with

$$f_1 = 1 - f_2, \; c_{12} = -\frac{c_{21}}{a_{21}}, \; a_{12} = \frac{1}{a_{21}} \quad , \tag{5.44}$$

$$\dot{f}_2 = c_{12} \frac{f_2(1 - f_2)}{1 + (a_{12} - 1) f_2} \quad .$$

To make the analysis as general as possible let us introduce the dimensionless time

$$\tau = \frac{t - t_h}{T_s} \quad ,$$

where t_h is the time point at which a half of the market is penetrated (i.e. for $t = t_h$ or $\tau = 0$, $f_1 = f_2 = 0.5$) and let us choose T_s in such a way that

$$\left.\frac{df_2(\tau)}{d\tau}\right|_{\tau = 0} = \frac{\ell n 9}{2} \tag{5.45}$$

in all cases. Obviously

$$\frac{df_2}{d\tau} = T_s \frac{df_2}{dt} = T_s c_{12} \frac{f_2(1 - f_2)}{1 + (a_{12} - 1)f_2} \tag{5.46}$$

and for $\tau = 0$ and $f_2 = 0.5$

$$\left.\frac{df_2}{d\tau}\right|_{\tau = 0} = \frac{T_s c_{12}}{2(1 + a_{12})} \quad . \tag{5.47}$$

Comparing the right-hand sides of (5.45) and (5.47) we have

$$T_s = \frac{a_{12} + 1}{c_{12}} \cdot \ell n 9 = \frac{\alpha_1 + \alpha_2}{c_1 - c_2 + (\alpha_1 - \alpha_2)\rho} \cdot \ell n 9 \quad . \tag{5.48}$$

Notice that for $a_{12} = 1$ (i.e. $\alpha_1 = \alpha_2$) T_s is the take-over time defined in the Fisher-Pry model as the time period required to transfer from $f_2 = 0.1$ to $f_2 = 0.9$ (see Section 2).

From (5.46) it is seen that after this normalization the differential equation describing the evolution of market shares gets the form

$$\frac{df_2}{d\tau} = (a_{12} + 1)\ \ell n 9 \cdot \frac{f_2(1 - f_2)}{1 + (a_{12} - 1)f_2} \tag{5.49}$$

and has to be solved for the initial condition $f_2 = f_1 = 0.5$ for $\tau = 0$. For any τ (positive or negative) the solution can be obtained using the subroutine PENETR with the following values of its formal parameters:

$$n = 2 \quad ;$$

$$\tilde{c}_1 = -(a_{12} + 1)\ell n 9 \quad ;$$

$$\tilde{c}_2 = 0 \quad ;$$

$$\tilde{\alpha}_1 = \frac{\alpha_1}{\alpha_2}$$

$$\tilde{\alpha}_2 = 1 \quad ;$$

$$f_{10} = 0.5 \quad ;$$

$$f_{20} = 0.5 \quad ;$$

$$t = \tau \quad .$$

The result of this calculation is given in Figures 12 and 13.

From Figure 13 it can be seen that within the take-over time the plot $\ell n(f_2/(1 - f_2))$ can well be approximated by a straight line for a rather large range of ratios $a_{12} = \alpha_1/\alpha_2$. This explains why the empirical Fisher-Pry model is able to describe so many practical cases even when the specific investments are different. It also conforms with the observation of C. Marchetti that the parameter c_{21} of the Fisher-Pry model can be determined from the present trend of the market penetration if the new technology reaches a nonnegligible part of the market, say approximately 10%.

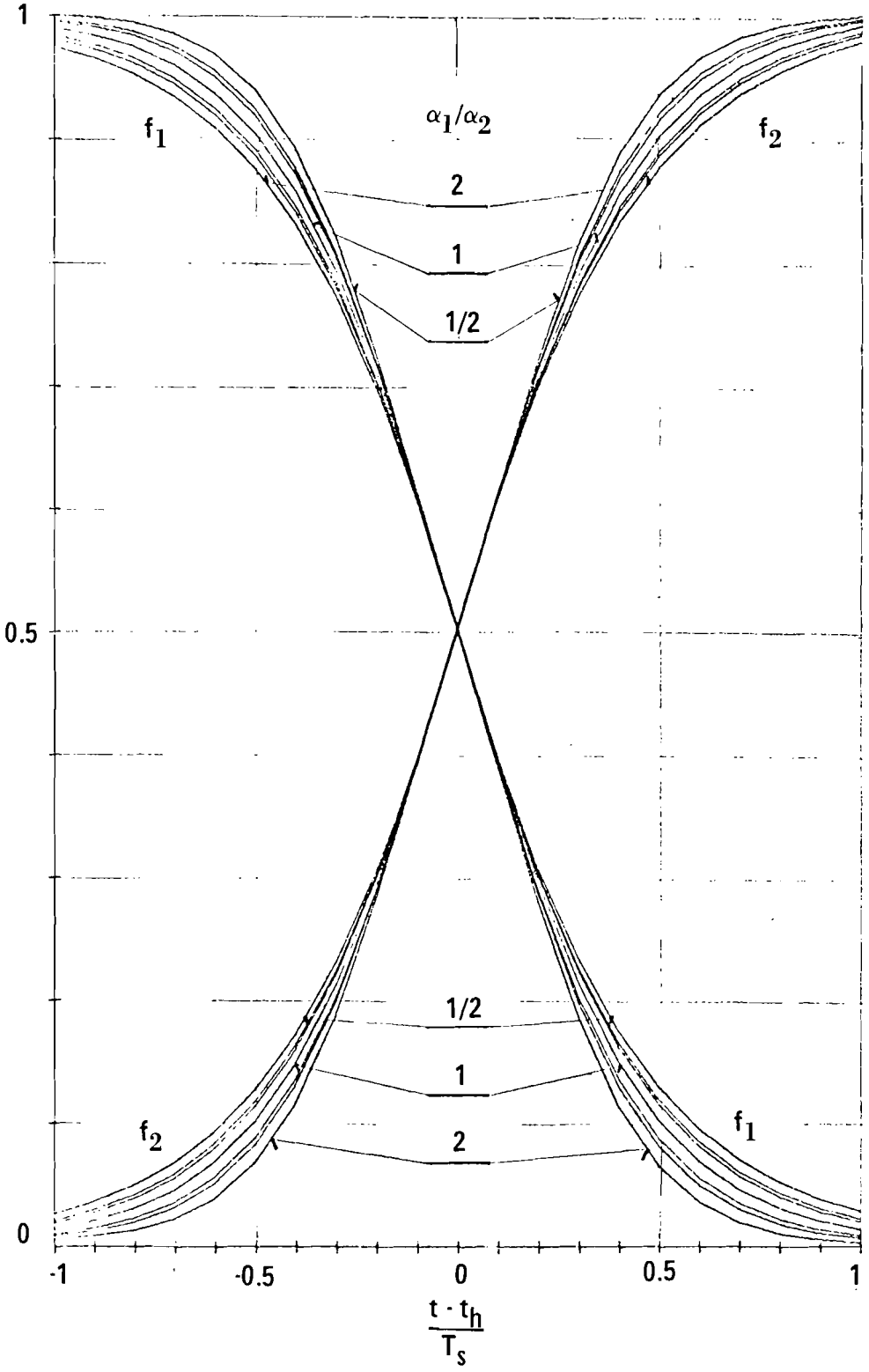


Figure 12. Two-dimensional competition. Sensitivity analysis with respect to different specific investments: $\alpha_1/\alpha_2 = 1/2, 2/3, 3/4, 1, 4/3, 3/2, 2$.

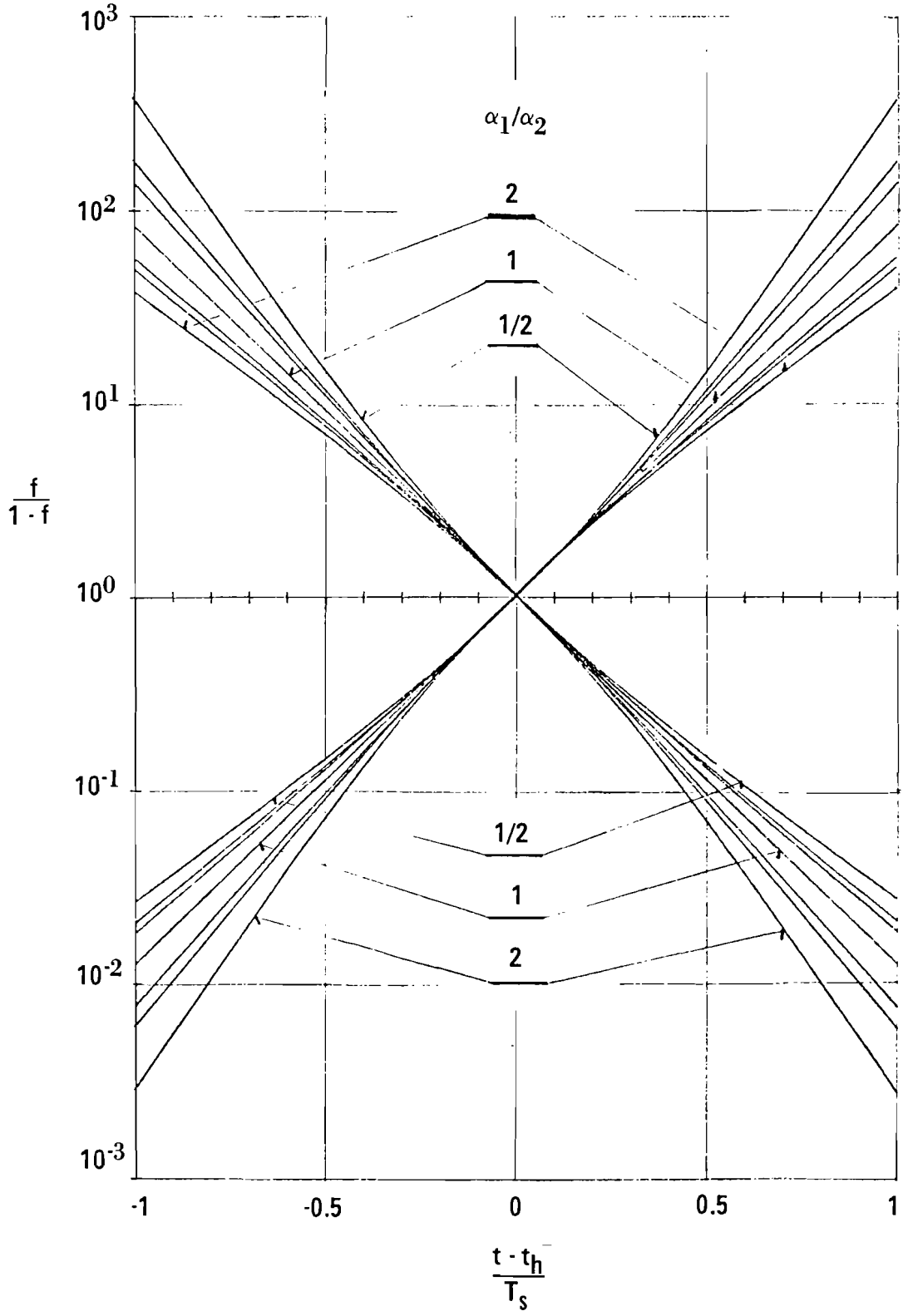


Figure 13. Two-dimensional competition. Sensitivity analysis with respect to different specific investments: $\alpha_1/\alpha_2 = 1/2, 2/3, 3/4, 1, 4/3, 3/2, 2$.

6. PROBABILISTIC MODEL AND ESTIMATION OF PARAMETERS FROM HISTORICAL DATA

To be able to use the model developed in the previous section for forecasting the future course of the substitution process we have to know the values of its parameters. In principle, there are two ways of obtaining these values: (a) the economic evaluation of the technologies entering the process, (b) the estimation of the parameters from known historical data. Each of these two possibilities has its advantages and drawbacks.

In some cases it may be very difficult to determine the model parameters using the former approach. For instance: How to calculate the correction Δp_i (see Section 4 for definition) by which the specific production costs c_i have to be reduced in the case when the competing commodities satisfy the given need in a qualitatively different way? Or, what is the precise meaning of the specific investment α_i in the case of wood as primary energy source and how this value can be calculated? However, there is no other way of proceeding when the substitution process has not started yet and no historical data are available. All what we can practically do in this case is to pick up one or more sets of possible values of model parameters according to our subjective judgment and to apply the model for these scenarios.

If the substitution process studied is already running in reality and its observation in the past is available, then the parameters -- whatever their interpretation is -- can be estimated on the basis of this past experience. How to extract the information about the unknown parameters from the historical data is the main question which will be studied in this section.

A widely used approach to parameter estimation is the so-called curve fitting. The parameters of the model are chosen in such a way that the output of the model is as close as possible to the known true data. When we try to formalize this ad hoc

approach mathematically several questions arise. The first one is: What curves have to be fitted? In our case we have many possibilities, for instance,

$$f_i(t), \quad \forall i; \quad \text{or } \ln \frac{f_i(t)}{1 - f_i(t)}, \quad \forall i; \quad \text{or } \ln \frac{f_i(t)}{f_j(t)}, \quad \forall i, j \neq i.$$

Which one of these (or other) possibilities has to be chosen to obtain the "most reliable" estimates?

A sound, intuitive, feeling says that the more data can be used the better estimates of the model parameters can be obtained. However, if the number of data is greater than the number of unknown parameters, a perfect fit can never be reached in general and the second question arises naturally: What significance has to be assigned to different errors and how to express it quantitatively?

To be able to answer these and similar questions in a consistent way we have to go deeper into the stochastic nature of the process and to build a probabilistic model.

To make this text accessible also to those readers who are not specialists in probability theory, we shall concentrate our exposition mainly on the underlying "philosophy" and practical results. The technical details of a mathematical character, which are not necessary for general understanding, will be stated in the main text without formal proofs. Full proofs (some of them are far from trivial) can be found in Appendix A.

Our approach to estimation and forecasting adopted throughout the rest of this paper is purely Bayesian. The substitution process we are studying is nonstationary in its nature: the situation which occurred in the past can never be repeated in the future. The set of historical data we have at our disposal, is one realization of one nonstationary stochastic process. Therefore the concept of probability cannot be based on frequency considerations. In the Bayesian view the probability is understood as a measure of belief and the probability distribution reflects

the uncertainty of the relationship between us (you or the author) and the external world. In general, different people may have different probability distributions for the same phenomenon depending on the information or knowledge they have. Because of this subjective feature one often speaks of the subjective probability approach. However, to avoid misunderstanding it should be stressed right at the beginning that using this approach it is possible to eliminate the prior subjective opinion about the possible values of unknown quantities and to base our judgment only on objective data. This will be shown in detail later on.

The mathematical system that is called Bayesian statistics, compared to other approaches to statistics, is the only one which is fully consistent and logically closed. Besides this mathematical beauty two pragmatic arguments speak in its favor. It is based on sound principles and works in practice, as we shall be able to show. An excellent explanation of the logical foundation of the probability theory with the interpretation outlined above has been given by De Finetti [13]. Very good textbooks are De Groot's [14] and Raiffa and Schlaifer's [15].

To those readers who are not familiar with the present state of mathematical statistics, it may seem that we devote too much space to the justification of the Bayesian approach. They are recommended to read a short but pithy talk given by Lindley at the conference on Directions for Mathematical Statistics [16].

The Bayesian position, from which essentially everything follows, is that all uncertain quantities -- including the unknown parameters -- are, before they are observed, random: that is, have a probability structure and a probability distribution can be assigned to them. This is, actually, no assumption; it can be proved on the basis of a few simple and natural axioms. The act of observation changes the status of the quantity observed from a random quantity to a number. If the quantities we are interested in, like parameters of our model, cannot be

observed directly but only through other related quantities, like the output of the model, they remain uncertain also after this observation; however, their probability distribution is changed -- their uncertainty is decreased. From this Bayesian point of view the parameter estimation means finding the probability distribution for the unknown parameters conditional on the observed data. Any single value, which is taken as an "estimate" is nothing more than some characteristic of this conditional distribution. It may be, for instance, the point at which the probability density reaches its maximum.

To be able to solve our practical problem only a few basic rules of the general probability theory have to be recalled.

Let a,b,c,... be random quantities or sets of such quantities and let $p(a|b)$ denote the probability density of a conditional on b. Then the following relations hold:

$$p(a,b|c) = p(a|b,c)p(b|c) \tag{6.1}$$

$$p(a|c) = \int p(a,b|c)\, db \quad , \tag{6.2}$$

where the integral in (6.2) is taken over all possible values of b. When (6.1) holds, the following relation must also hold:

$$p(a,b|c) = p(b|a,c)p(a|c) \quad . \tag{6.3}$$

From (6.3)

$$p(b|a,c) = \frac{p(a,b|c)}{p(a|c)}$$

and using (6.1) and (6.2) we have

$$p(b|a,c) = \frac{p(a|b,c)p(b|c)}{\int p(a|b,c)p(b|c)\, db} \quad . \tag{6.4}$$

This is the famous Bayes rule which makes it possible to determine $p(b|a,c)$ when $p(a|b,c)$ and $p(b|c)$ are known.

A successive application of (6.1) gives an other useful tool sometimes called the chain rule for probability densities.

$$p(x_1,x_2,\ldots,x_i,\ldots,x_N) = \prod_{i=1}^{N} p(x_i|x_{i-1},x_{i-2},\ldots,x_1) \quad . \tag{6.5}$$

Later on we shall also need the rule according to which the probability density $p(x_1,x_2,\ldots,x_n)$ can be recalculated into the probability density $p(y_1,y_2,\ldots,y_n)$ when the random quantities $\{x_1,x_2,\ldots,x_n\}$ and $\{y_1,y_2,\ldots,y_n\}$ are related by a regular (one-to-one) deterministic transformation. Let

$$p(x_1,x_2,\ldots,x_n) = g(x_1,x_2,\ldots,x_n) \quad ,$$

$$x_i = \xi_i(y_1,y_2,\ldots,y_n) \quad ; \quad i = 1,2,\ldots,n \quad , \tag{6.6}$$

then

$$p(y_1,y_2,\ldots,y_n) = g(\xi_1,\xi_2,\ldots,\xi_n) \cdot |J| \quad , \tag{6.7}$$

where $|J|$ means the absolute value of the determinant J (Jacobian) of the transformation (6.6).

$$J = \frac{\mathcal{D}(\xi_1,\xi_2,\ldots,\xi_n)}{\mathcal{D}(y_1,y_2,\ldots,y_n)} = \begin{vmatrix} \frac{\partial\xi_1}{\partial y_1}, \frac{\partial\xi_1}{\partial y_2}, \ldots, \frac{\partial\xi_1}{\partial y_n} \\ \frac{\partial\xi_2}{\partial y_1}, \frac{\partial\xi_2}{\partial y_2}, \ldots, \frac{\partial\xi_2}{\partial y_n} \\ \vdots \\ \frac{\partial\xi_n}{\partial y_1}, \frac{\partial\xi_n}{\partial y_2}, \ldots, \frac{\partial\xi_n}{\partial y_n} \end{vmatrix} \quad . \tag{6.8}$$

Now, let us go closer to our estimation problem using these methodological tools.

Let K be the finite set of unknown parameters we would like to know and let D be the set of data we have at our disposal.

Then the probability distribution we are interested in is $p(K|D)$. Let ν be the dimension of K (the number of unknown parameters) and S_K^ν be the space of all possible values of K which have to be considered. Applying the Bayes rule (6.4) we have

$$p(K|D) = \frac{p(D|K)p(K)}{\int_{S_K^\nu} p(D|K)p(K)\,dK} \quad , \tag{6.9}$$

where $p(K)$ is the prior (subjective) probability density which we -- as users of this tool -- have to assign to the unknown parameters before the observed data are incorporated into our knowledge. The probability density of the set of the observed quantities D given the parameters K has to be known. This is one of the reasons for which we have to build the probabilistic model of the process. We shall come to this point after this general introduction where we want to explain the basic philosophy of Bayesian estimation.

The operation (6.9) can be understood as the correction of our prior subjective probability distribution for K by objective data. A classical objection to this Bayesian estimation is that when the prior distribution $p(K)$ is wrong in the sense that it prefers other than true values of K, then the resulting $p(K|D)$ is incorrect or at least biased. This objection is fully justified but should be oriented not against the Bayesian statistics but against the user. Mathematics provides us with a logical and consistent system of reasoning but it cannot correct our mistakes. The prior distribution $p(K)$ is a model of our prior uncertainty. Like any other mathematical model, for mathematics it is an input. If the input is wrong the output is also wrong in general. The system of axioms, on which mathematics operates, and mathematical models are the only connection between mathematics and the true world.

If the reader accepts this explanation he has the right to ask the following question: How should we choose the model $p(K)$

to be sure that no subjective mistake is introduced into the calculation?

The author does not know any practical situation when absolutely no prior information is available. For instance, in our case of market penetration the parameter a_{ir} means the ratio of specific investments, $a_{ir} = \alpha_i/\alpha_r$, and therefore cannot be negative. But we also a priori know that it cannot be larger than say 10^{12}. Similarly, we can assume with certainty that the parameter c_{ir} lies within the interval $-10^{30} < c_{ir} < 10^{30}$. In practice we can choose any prior distribution $p(K)$ which is very "flat", but if we want to be extremely "objective" we may choose the uniform distribution

$$p(K) = \varepsilon \quad \text{for } K \in S_K^\nu$$

$$p(K) = 0 \quad \text{for } K \notin S_K^\nu \quad ,$$

where

$$\varepsilon = \frac{1}{\int\limits_{S_K^\nu} dS} \quad . \tag{6.10}$$

In that case (6.6) gives

$$p(K|D) = \frac{p(D|K)}{\int\limits_{S_K^\nu} p(D|K)dK} \quad \text{for} \quad K \in S_K^\nu \tag{6.11}$$

$$p(K|D) = 0 \quad \text{for} \quad K \notin S_K^\nu \quad .$$

Strictly speaking, the space S_K^ν can be as large as we want but finite. It is not possible to distribute uniformly one unit (of our belief) on an infinite countable set of intervals. If S_K^ν is growing to infinity then ε, defined by (6.10), tends to zero and the right-hand side of (6.9) becomes undetermined.

For technical reasons it is usually much more convenient to

operate with probability density functions which are defined by a unique formula on the whole euclidean space R^ν rather than on its subset $S_K^\nu \subset R^\nu$, like (6.11). (This is actually the reason why the normal distribution is so widely used also in cases when it is known that the random quantity may lie only within a finite interval.) This situation can be reached also for the conditional distribution (6.11) if the integral

$$\int_{R^\nu} p(D|K)\, dK$$

exists. Under this condition (6.11) can be written

$$p(K|D) = \frac{p(D|K)}{\int_{R^\nu} p(D|K)\, dK - \int_{R^\nu - S_K^\nu} p(D|K)\, dK}$$

and for $S_K^\nu \to R^\nu$ we obtain the simple relation

$$p(K|D) = \frac{1}{\kappa(D)} \cdot p(D|K) \quad , \tag{6.12}$$

which holds for all $K \in R^\nu$ and where $\kappa^{-1}(D)$ is the normalizing factor

$$\kappa(D) = \int p(D|K) dK \quad , \tag{6.13}$$

which does not depend on the unknown parameters K.

The probability density $p(D|K)$, considered -- for given data D -- as a function of unknown parameters K, is called the <u>likelihood function</u>

$$L(K) = p(D|K) \quad . \tag{6.14}$$

If we are interested in the point in which the probability density $p(K|D)$ reaches its maximum we have to find such a set of

possible values of unknown parameters $\hat{K}$ for which

$$L(\hat{K}) \geq L(K) \quad . \tag{6.15}$$

The point $\hat{K}$ is called the maximum likelihood (ML) estimate. When $\hat{K}$ is expressed as a function of observed data D, $\hat{K} = \hat{K}(D)$, then one speaks about a maximum likelihood (ML) estimator.

A classical (non-Bayesian) statistician may say that this result is nothing else than the well-known and nowadays almost generally accepted method of parameter estimation in non-Bayesian ("objective") statistics. He is right--formally. There is a great difference not only in understanding and interpretation of this result, but also in practical application: (a) The maximum likelihood method cannot be derived within non-Bayesian statistics. It can be only proposed -- as an inductive step of reasoning (opposite to logically deductive) -- and its properties can be investigated ex post. (b) All general results concerning the properties of ML estimates, known i non-Bayesian statistics, are of asymptotic character, i.e. they apply for a large number of samples which are independently drawn from the same distribution. Accepting the Bayesian point of view we know precisely what we are doing for any sample size. (c) The choice of one point from the whole distribution (6.12) is, in fact, a decision problem and as such has to be handled. Depending on the final goals we are pursuing, some other points may be more suitable than the maximum of the probability density. (d) In many cases -- like forecasting -- we are actually not interested in point estimates. Parameter estimation is usually only one step in the solution of the whole problem. In such cases it is possible -- and in general also necessary -- to consider all possible values of the parameters, i.e. to operate with the whole distribution. This will be clearly seen in the next section where we shall deal with forecasting.

Let us now reconsider the question which arised at the beginning of this section in connection with curve fitting, namely

"What curves have to be fitted?", and let us show that it is irrelevant. The question is equivalent to the question whether the parameter estimation is influenced by regular (one-to-one) transformation of the data set used. Let $\tilde{D}$ be the data set obtained from the original one D by a regular transformation

$$\tilde{D} = F(D)$$

$$D = \phi(\tilde{D}) \tag{6.16}$$

and let

$$p(D|K) = g(D,K) \quad .$$

Then, according to (6.7), the conditional probability density for the new data set $\tilde{D}$ given K is

$$p(\tilde{D}|K) = g(\phi(\tilde{D}),K) \cdot |J(\tilde{D})| \quad , \tag{6.17}$$

where $J(\tilde{D})$ is the Jacobian

$$J(\tilde{D}) = \frac{\mathcal{D}(\phi)}{\mathcal{D}(\tilde{D})} \quad . \tag{6.18}$$

The transformation (6.16), and consequently also the Jacobian (6.18) cannot depend on the unknown parameters, otherwise, not knowing the parameters, we would not be able to recalculate the data and use them as input for our estimation problem. Using (6.17) and (6.12), for $\tilde{D}$ instead of D, we obtain

$$p(K|\tilde{D}) = \frac{g(\phi(\tilde{D}),K)}{\int g(\phi(\tilde{D}),K)\, dK} = p(K|D) \quad , \tag{6.19}$$

which proves that not only the ML estimates but the entire probability distribution for K is invariant with respect to one-to-one transformation of the data set D.

It is also possible to show that the ML estimates are invariant with respect to one-to-one transformation of unknown parameters K, i.e. that

$$\hat{\phi}(K) = \phi(\hat{K}) \quad .$$

To proceed towards our practical problem we have to show how the probability density $p(D|K)$ can be calculated and what we mean by a probabilistic model. The processes we are going to consider are time oriented. Every observation we can make is related to some time point or time interval. If we want to describe and analyze the relations between these quantities we have to distinguish them; we have to specify the set D in more detail. Let $y_{(k)}$ be the quantity, in general a vector, which can be observed at time t_k. We shall call $y_{(k)}$ the <u>output of the process</u> at time t_k. Let us order the time indexing in such a way that $t_{k-1} < t_k < t_{k+1}$, i.e. the output $y_{(k-1)}$ precedes the output $y_{(k)}$. Let $y_{(1)}$ be the first and $y_{(N)}$ the last output the observations of which are available. To simplify the writing we shall introduce the following notation for sets of outputs

$$y_{(i)}^{(j)} = \left\{ y_{(j)}, y_{(j-1)}, \ldots, y_{(i+1)}, y_{(i)} \right\} = \left\{ y_{(j)}, y_{(i)}^{(j-1)} \right\}; \quad j \geq i \; . \tag{6.20}$$

For $j < i$ the set (6.20) is empty. With this notation the data set D is

$$D = y_{(1)}^{(N)}$$

and the probability density which has to be known is

$$p(D|K) = p\left(y_{(1)}^{(N)} | K \right) \quad .$$

Using the chain rule (6.5) we can expand this density in the following way

$$p\left(y_{(1)}^{(N)} | K \right) = \prod_{k=1}^{N} p\left(y_{(k)} | y_{(1)}^{(k-1)}, K \right) \quad . \tag{6.21}$$

To keep the contact with reality we have to understand what the particular factor in this product, namely

$$p\left(y_{(k)} | y_{(1)}^{(k-1)}, K \right) \quad , \tag{6.22}$$

physically means. The probability density (6.22) describes the probabilistic transformation between the past history of the process $y_{(1)}^{(k-1)}$ and the next output $y_{(k)}$. It is the probabilistic law of the evolution of the process. Notice that the system

$$\left\{ p\left(y_{(k)} \mid y_{(1)}^{(k-1)}, K \right) : k = 1,2,..... \right\} \tag{6.23}$$

is the most general description of the process from the viewpoint of the outer observer that makes it possible to determine (using two basic rules (6.1) and (6.2)) any finite dimensional probability density for any combination of quantities which can be observed on the process. K is the finite set of parameters which are unknown in this system of functions.

By a probabilistic model of the process we mean any mathematical description of the process which defines the conditional probability density (6.22) for any k up to a finite set of parameters K. In the sequel we shall develop such a model for our case of market penetration by new technologies.

6.1 Probabilistic Model

Approaching any practical modelling problem we have to specify, first of all, to what goal the model has to serve. In our case this goal is: (1) to explain and to identify, i.e. to describe quantitatively, the past evolution of market shares $f_i(t)$, $i = 1,2,...,n$; (2) to forecast the future evolution of market shares.

The second question which must be cleared is: What data are available to identify the process? In this study we shall assume that only the market shares at discrete time points (not necessarily equally spaced) are available.

The market shares fulfill the deterministic relation

$$\sum_{i=1}^{n} f_i(t) = 1$$

and therefore one of them, say

$$f_r(t) = 1 - \sum_{i \neq r} f_i(t) \quad , \tag{6.24}$$

can be omitted in the probabilistic model. When $f_i(t_k)$, $\forall i \neq r$, are known, the value $f_r(t_k)$ does not bring any new information and when we are able to forecast $f_i(t)$, $\forall i \neq r$, we are also able to forecast $f_r(t)$. Hence, the output of the process, we are studying, is the (n-1)-dimensional vector

$$y_{(k)} = \{f_i(t_k) : \forall i \neq r\} \quad . \tag{6.25}$$

In algebraic expressions $y_{(k)}$ will mean a column vector the (n-1) components of which are ordered in an arbitrary but fixed way.

Our modelling effort is to find the conditional probability density

$$p\left(y_{(k)} \mid y_{(1)}^{(k-1)}, K\right) \quad . \tag{6.26}$$

As we cannot assume the prior knowledge of any parameter of this distribution, K will be the set of all parameters. This means that we have -- on the basis of sound and realistic assumptions -- to find only the structure of this function. Pursuing this aim we shall start again with the equation (3.1)

$$\alpha_i [P_i(t+\Delta t) - P_i(t)] = \int_t^{t+\Delta t} P_i(t)[p(t) - c_i]dt + Q_i(t,\Delta t) + \Delta_i(t,\Delta t) \tag{6.27}$$

which formed the basis of our deterministic model. However, building the probabilistic model we have to consider also the stochastic terms. The equation (6.27) can be rewritten as follows:

$$\int_t^{t+\Delta t} \{\alpha_i \dot{P}_i(t) - P_i(t)[p(t) - c_i] - \delta q_i(t)\}dt = 0 \quad , \tag{6.28}$$

where $\delta q_i(t)$ now means a stochastic process for which

$$\int_t^{t+\Delta t} \delta q_i(t)dt = Q_i(t,\Delta t) + \Delta_i(t,\Delta t) \quad . \tag{6.29}$$

From (6.28) the stochastic differential equation obtained is:

$$\alpha_i \dot{P}_i(t) = P_i(t)[p(t) - c_i] + \delta q_i(t) \quad . \tag{6.30}$$

We shall again assume that $\delta q_i(t)$ is zero but only in the mean.

Assumption 1

$$E[\delta q_i(t)] = 0 \quad , \quad \forall i \quad . \tag{6.31}$$

This means that we admit stochastic fluctuations around zero both for the external capital flow and for the capital reserves. Notice that the process $\delta q_i(t)$ is not necessarily white (actually no real continuous process can be white). All that we claim until now is (6.29) and (6.31).

It is not unrealistic to assume that the standard deviation of the stochastic fluctuations $\delta q_i(t)$ around the zero mean value is proportional to the instantaneous production $P_i(t)$. As $P_i(t)$ can be only positive we can write

$$q_i(t) = P_i(t) \cdot \delta c_i(t) \quad . \tag{6.32}$$

This transformation of the stochastic process $q_i(t)$ seems to be reasonable except at the very beginning when $P_i(t)$ is close to zero and the new technology needs some external capital input to be able to start the production. The assumption we are discussing is only a part of an assumption which will be made later on in a more formal and precise way.

Substitution of (6.32) into (6.30) gives

$$\alpha_i \dot{P}_i(t) = P_i(t)[p(t) - c_i + \delta c_i(t)] \quad . \tag{6.33}$$

Notice also that stochastic fluctuations of c_i can be incorporated into $\delta c_i(t)$. This is, actually, the reason for the notation used.

The market price $p(t)$ can be eliminated in a similar way as in the deterministic case in Section 5 if the following assumption is accepted.

Assumption 2

$$\frac{\dot{P}(t)}{P(t)} = \rho + \delta\rho(t) \tag{6.34}$$

$$E\delta\rho(t) = 0 \ , \tag{6.35}$$

where ρ is the growth rate of the total production $P(t)$ of all competing technologies and $\delta\rho(t)$ is the stochastic fluctuation around the constant ρ. Similarly to (5.2) we have

$$\frac{\dot{P}_i(t)}{P_i(t)} = \frac{\dot{f}_i(t)}{f_i(t)} + \frac{\dot{P}(t)}{P(t)} = \frac{\dot{f}_i(t)}{f_i(t)} + \rho + \delta\rho(t) \ ,$$

and the following stochastic analogy of (5.4) is obtained.

$$\alpha_i \frac{\dot{f}_i}{f_i} + c_i + \alpha_i\rho - \left(\alpha_r \frac{\dot{f}_r}{f_r} + c_r + \alpha_r\rho\right) = \delta c_i - \delta c_r + (\alpha_i - \alpha_r)\delta\rho$$

$$\frac{\dot{f}_i}{f_i} - a_{ri}\frac{\dot{f}_r}{f_r} - c_{ri} = \delta e_{ri} \ , \tag{6.36}$$

where a_{ri} and c_{ri} are the parameters, defined by (5.11) and (5.12), and

$$\delta e_{ri}(t) = \frac{1}{\alpha_i}(\delta c_i - \delta c_r) + \left(1 - \frac{\alpha_r}{\alpha_i}\right)\delta\rho \ . \tag{6.37}$$

Integration of the equation (6.36) over the time interval

(t_{k-1}, t_k) gives

$$\ell nf_{(k)i} - \ell nf_{(k-1)i} - a_{ri}(\ell nf_{(k)r} - \ell nf_{(k-1)r}) - c_{ri}(t_k - t_{k-1}) = e_{(k)ri} \quad , \tag{6.38}$$

where

$$f_{(k)i} = f_i(t_k) \tag{6.39}$$

$$e_{(k)ri} = \int_{t_{k-1}}^{t_k} \delta e_{ri}(t)\,dt \quad . \tag{6.40}$$

As the mean value of $\delta e_{ri}(t)$ is zero for any t the integral (6.40) also has this property.

$$E e_{(k)ri} = 0 \quad , \qquad \forall k,i \quad . \tag{6.41}$$

Let us introduce the following notation for the column vector of all $e_{(k)ri}$, $i \neq r$.

$$e_{(k)} = col[e_{(k)ri} : \forall i \neq r] \quad . \tag{6.42}$$

In correspondence with (6.25) and (6.24) we also have

$$y_{(k)} = col[f_{(k)i}\,, \forall i \neq r] \tag{6.43}$$

$$f_{(k)r} = 1 - \sum_{i \neq r} f_{(k)i} = 1 - \sum_{j=1}^{n-1} y_{(k)j} \quad . \tag{6.44}$$

For later use we shall also introduce the (n-1) - vectors of parameters

$$c = col[c_{ri} : \forall i \neq r] \tag{6.45}$$

$$a = col[a_{ri} : \forall i \neq r] \quad . \tag{6.46}$$

For given parameters a, c and given $y_{(1)}$ the equation (6.38) together with (6.44) define a deterministic transformation between the stochastic processes $\{y_{(k)} : k = 2,3,\ldots\}$ and $\{e_{(k)} : k = 2,3,\ldots\}$. This means that if we are able to find a suitable conditional probability distribution $p(e_{(k)} | e_{(2)}^{(k-1)}, y_{(1)}, K)$ we shall also be able to find the density $p(y_{(k)} | y_{(1)}^{(k-1)}, K)$ we are looking for. This is the line we shall follow now.

If the sampling interval $(t_k - t_{k-1})$ is large enough it is realistic to assume that the random variable $e_{(k)}$ is independent of the previous ones $e_{(2)}^{(k-1)}$ and also of $y_{(1)}$.

Assumption 3

$$p\left(e_{(k)} | e_{(2)}^{(k-1)}, y_{(1)}, K\right) = p(e_{(k)} | K) \quad . \tag{6.47}$$

This assumption means that the knowledge of $e_{(2)}^{(k-1)}$ and $y_{(1)}$ (or equivalently the knowledge of $y_{(1)}^{(k-1)}$) cannot bring any information about the possible value of $e_{(k)}$. Because of the deterministic relation between $y_{(2)}^{(k-1)}$ and $e_{(2)}^{(k-1)}$ (for given $y_{(1)}$ and parameters) the conditional part of (6.47) can be modified as follows.

$$p\left(e_{(k)} | y_{(1)}^{(k-1)}, K\right) = p\left(e_{(k)} | e_{(2)}^{(k-1)}, y_{(1)}, K\right) = p(e_{(k)} | K) . \tag{6.48}$$

As shown in Appendix A (Theorem 1) the transformation between $e_{(k)}$ and $y_{(k)}$ for given $y_{(k-1)}$ and for positive a_{ri}, $\forall i$, is one-to-one with the Jacobian

$$J_{ey} = \frac{\partial(e_{(k)})}{\partial(y_{(k)})} = \frac{f_{(k)r} + \sum_{i \neq r} a_{ri} f_{(k)i}}{\prod_{i=1}^{n} f_{(k)i}} \quad . \tag{6.49}$$

Hence, according to (6.7), we have

$$p(y_{(k)}|y_{(1)}^{(k-1)},K) = |J_{ey}|\cdot p(e_{(k)}|y_{(1)}^{(k)},K)$$
$$= |J_{ey}|\cdot p(e_{(k)}|K) \quad . \tag{6.50}$$

This shows that all that remains to be found is a suitable structure of the probability density $p(e_{(k)}|K)$ with a minimum number of unknown parameters.

We have no reasons to prefer positive or negative values of the random variable $e_{(k)}$. It means that the density $p(e_{(k)}|K)$ has to be symmetric. From the left-hand side of (6.38) it can be seen that the density should be defined over the whole range R^{n-1} of possible values $e_{(k)}$. These requirements are fulfilled by a multivariate normal distribution with zero mean. Making this choice of the form of the distribution we have to define the covariance matrix

$$R_{(k)} = E[e_{(k)}e_{(k)}^T]$$

for any k through a finite number of unknown constants.

If we want to consider also the cases when the samples $y_{(k)}$ are not equally spaced in time, i.e. the interval $(t_k - t_{k-1})$ may be different for different k, then the suitable structure for $R_{(k)}$ is

$$R_{(k)} = (t_k - t_{k-1})R \tag{6.51}$$

where R is an unknown but constant matrix. To show the relevance of this structure let us divide the time interval $(t_k - t_{k-1})$ into μ equal intervals

$$t_k - t_{k-1} = \mu\Delta t \quad .$$

From the definition of the random variable $e_{(k)ir}$ (6.40) we have

$$e_{(k)} = \int_{t_{k-1}}^{t_k} \delta e(t)\,dt = \sum_{m=1}^{\mu} \int_{t_{k-1}+(m-1)\Delta t}^{t_{k-1}+m\Delta t} \delta e(t)\,dt$$

$$e_{(k)} = \sum_{m=1}^{\mu} e_{(k)m} \quad , \tag{6.52}$$

where

$$e_{(k)m} = \int_{t_{k-1}+(m-1)\Delta t}^{t_{k-1}+m\Delta t} \delta e(t)\,dt \quad .$$

If the time interval Δt is still large enough that the random variables $\{e_{(k)m} : m + 1,2,\ldots,\mu\}$ can be assumed to be uncorrelated

$$E[e_{(k)m} e^T_{(k)j}] = 0 \quad , \text{ for } \quad j \neq m \quad ,$$

$$E[e_{(k)m} e^T_{(k)m}] = M \quad ,$$

we obtain using the relation (6.52)

$$R_{(k)} = \sum_{m=1}^{\mu} E e_{(k)m} e^T_{(k)m} = \mu M \quad .$$

This result shows that if we assume the independence (6.47) then the covariance matrix $R_{(k)}$ has to be proportional to the length of the sampling interval $(t_k - t_{k-1})$. This justifies the structure (6.51).

Summing up we can make the last assumption.

Assumption 4

$$p(e_{(k)}|K) = (2\pi)^{-\frac{(n-1)}{2}} \frac{|\Omega|^{\frac{1}{2}}}{(t_k - t_{k-1})^{\frac{n-1}{2}}} \exp\left\{-\tfrac{1}{2} e^T_{(k)} \frac{\Omega}{t_k - t_{k-1}} e_{(k)}\right\} \quad , \tag{6.53}$$

where we introduced for later convenience the precision matrix [14]

$$\Omega = R^{-1} \tag{6.54}$$

as the unknown parameter instead of R. Thus the full set of unknown parameters is

$$K = \{c,a,\Omega\} \quad . \tag{6.55}$$

To make the formulae more compact it is convenient to introduce the following n-vectors

$$f_{(k)} = \begin{bmatrix} y_{(k)} \\ f_{(k)r} \end{bmatrix} = col[f_{(k)i} : \forall i\,] \tag{6.56}$$

$$\tilde{a} = \begin{bmatrix} a \\ 1 \end{bmatrix} \quad , \tag{6.57}$$

the $[(n-1) \times n]$-matrix

$$A = [I_{n-1}, -a] \quad , \tag{6.58}$$

where I_{n-1} is an identity matrix of dimension (n-1), and also the following notation

$$\Pi(f_{(k)}) = \prod_{i=1}^{n} f_{(k)i} \tag{6.59}$$

$$T_{(k)} = t_k - t_{k-1} \tag{6.60}$$

$$x_{(k)i} = \left(\ell n\, f_{(k)i} - \ell n\, f_{(k-1)i} \right) = \ell n\, \frac{f_{(k)i}}{f_{(k-1)i}} \tag{6.61}$$

$$x_{(k)} = col[x_{(k)i} : \forall i] \quad . \tag{6.62}$$

With this notation the density (6.53) gets the form

$$p(e_{(k)}|K) = (2\pi)^{-\frac{n-1}{2}} \frac{|\Omega|^{\frac{1}{2}}}{T_{(k)}^{\frac{n-1}{2}}} \exp\left\{-\frac{1}{2T_{(k)}} e_{(k)}^T \Omega e_{(k)}\right\}, \tag{6.63}$$

the Jacobian (6.49) is

$$J_{ey} = \frac{f_{(k)}^T \tilde{a}}{\Pi(f_{(k)})} \quad , \tag{6.64}$$

and $\tilde{e}_{(k)}$ can be expressed from (6.38) as follows:

$$e_{(k)} = A x_{(k)} - c T_{(k)} \, . \tag{6.65}$$

Substitution into (6.50) gives the probability density which we are looking for.

$$p\left(y_{(k)}|y_{(1)}^{(k-1)},K\right) = (2\pi)^{-\frac{n-1}{2}} \frac{f_{(k)}^T \tilde{a}}{\Pi(f_{(k)})} \frac{|\Omega|^{\frac{1}{2}}}{T_{(k)}^{\frac{n-1}{2}}} \exp\left\{-\frac{1}{2T_{(k)}} e_{(k)}^T \Omega e_{(k)}\right\} . \tag{6.66}$$

Notice that $\tilde{e}_{(k)}$ and thereby the entire conditional density (6.66) for the output $y_{(k)}$ depends only on the preceding output $y_{(k-1)}$. Hence, the old history $y_{(1)}^{(k-2)}$ can be omitted in the condition part

$$p\left(y_{(k)}|y_{(1)}^{(k-1)},K\right) = p(y_{(k)}|y_{(k-1)},K) \quad , \tag{6.67}$$

and the output $y_{(k)}$ can be considered as an observable state of a nonlinear dynamic stochastic system.

The conditional probability density (6.66) is all that we need to be able to estimate the parameters and to forecast the future course of the process on the basis of known historical data.

6.2 Maximum Likelihood Estimates

As mentioned above, to estimate the unknown parameters $K = \{c,a,\Omega\}$ means finding the probability distribution for this parameter set conditioned on the known data

$$p\left(K|y_{(1)}^{(N)}\right) = p\left(c,a,\Omega|y_{(1)}^{(N)}\right) \quad . \tag{6.68}$$

The dynamic probabilistic model, developed in the previous subsection, makes it possible to calculate

$$p\left(y_{(2)}^{(N)}|y_{(1)},K\right) = \prod_{k=2}^{N} p\left(y_{(k)}|y_{(1)}^{(k-1)},K\right) \quad , \tag{6.69}$$

where $p\left(y_{(k)}|y_{(1)}^{(k-1)},K\right)$ is the density (6.66). Using the Bayes rule (6.4) we can determine the a posteriori probability density (6.68) in the following way.

$$p\left(K|y_{(1)}^{(N)}\right) = \frac{p\left(y_{(2)}^{(N)}|y_{(1)},K\right) \cdot p(K|y_{(1)})}{\int p\left(y_{(2)}^{(N)}|y_{(1)},K\right) \cdot p(K|y_{(1)})dK} \quad . \tag{6.70}$$

Considering that the single observation of the first output $y_{(1)}$ does not bring any information about the unknown parameters K, we can write

$$p(K|y_{(1)}) = p(K) \quad , \tag{6.71}$$

where

$$p(K) = p(c,a,\Omega)$$

is the prior distribution for the unknown parameters.

Using similar arguments as in the general introductory part of this section we obtain for the limit case of a very flat prior subjective probability distribution

$$p\left(K|y_{(1)}^{(N)}\right) = \frac{1}{\kappa\left(y_{(1)}^{(N)}\right)} \cdot L\left(c,a,\Omega,y_{(1)}^{(N)}\right) , \qquad (6.72)$$

where

$$L\left(c,a,\Omega,y_{(1)}^{(N)}\right) = p\left(y_{(2)}^{(N)}|y_{(1)},K\right) \qquad (6.73)$$

is the likelihood function for our estimation problem and

$$\kappa\left(y_{(1)}^{(N)}\right) = \int_{S_c}\int_{S_a}\int_{S_\Omega} L\left(c,a,\Omega,y_{(1)}^{(N)}\right) dc \cdot da \cdot d\Omega . \qquad (6.74)$$

As the components of the vector a (i.e., $a_{ri} = \alpha_r/\alpha_i, i\neq r$) can be only positive the space S_a is a space of all (n-1)-vectors with positive components. S_Ω is the space of all positive definite matrices of dimension (n-1) × (n-1) and S_c is the euclidean (n-1)-dimensional space $R^{(n-1)}$. It is assumed, of course, that N is large enough so that the integral (6.74) exists.

The likelihood function (6.73) is obtained as the product (6.69) of probability densities (6.66). From the properties of the trace of matrix expressions

$$\operatorname{tr} B + \operatorname{tr} C = \operatorname{tr}(B+C)$$

$$\operatorname{tr}(B\,C) = \operatorname{tr}(C\,B) ,$$

it follows that

$$\sum_{k=2}^{N} \frac{1}{T_{(k)}} e_{(k)}^T \Omega e_{(k)} = \sum_{k=2}^{N} \operatorname{tr}\left[\Omega e_{(k)} \frac{1}{T_{(k)}} e_{(k)}^T\right] = \operatorname{tr}\left[\Omega \sum_{k=2}^{N} e_{(k)} \frac{1}{T_{(k)}} e_{(k)}^T\right] . \qquad (6.75)$$

When the right-hand side of (6.65) is used to express the vectors $e_{(k)}$ we get

$$\sum_{k=2}^{N} \frac{1}{T_{(k)}} e_{(k)}^T \Omega e_{(k)} = A\phi_{(N)}A^T - Am_{(N)}c^T - c\,m_{(N)}^T A^T + c\tau_{(N)}c^T , \qquad (6.76)$$

where

$$\phi_{(N)} = \sum_{k=2}^{N} \frac{1}{T_{(k)}} x_{(k)} x_{(k)}^{T} \tag{6.77}$$

$$m_{(N)} = \sum_{k=2}^{N} x_{(k)} \tag{6.78}$$

$$m_{(N)i} = \sum_{k=2}^{N} \ell n \frac{f_{(k)i}}{f_{(k-1)i}} = \ell n \frac{f_{(N)i}}{f_{(1)i}} \tag{6.79}$$

$$\tau_{(N)} = \sum_{k=2}^{N} T_{(k)} = t_{(N)} - t_{(1)} \quad . \tag{6.80}$$

Using (6.76) the likelihood function can be brought into the following form:

$$L\left(c,a,\Omega,y_{(1)}^{(N)}\right) = p\left(y_{(2)}^{(N)} \mid y_{(1)},c,a,\Omega\right)$$

$$= \frac{1}{(2\pi)^{\frac{N(n-1)}{2}} \left(\prod_{k=2}^{N} T_{(k)}\right)^{\frac{n-1}{2}} \cdot \prod_{k=2}^{N} \Pi(f_{(k)})}$$

$$\times \prod_{k=2}^{N} \left(f_{(k)}^{T} \tilde{a}\right) \cdot |\Omega|^{\frac{N-1}{2}} \exp\left\{-\tfrac{1}{2} \operatorname{tr}\left[\Omega\left(A\phi_{(N)} A^{T} - 2Am_{(N)}c^{T} + c\tau_{(N)}c^{T}\right)\right]\right\} . \tag{6.81}$$

As the first factor, not depending on the unknown parameters, cancels in (6.72) it is possible and more convenient to operate with the modified likelihood function

$$\tilde{L}\left(c,a,\Omega,y_{(1)}^{(N)}\right) = \prod_{k=2}^{N} \left(f_{(k)}^{T} \tilde{a}\right) \cdot |\Omega|^{\frac{N-1}{2}} \exp\left\{-\tfrac{1}{2}\operatorname{tr}\left[\Omega\left(A\phi_{(N)}A^{T} - 2Am_{(N)}c^{T} + c\tau_{(N)}c^{T}\right)\right]\right\} \tag{6.82}$$

$$p\left(c,a,\Omega|y_{(1)}^{(N)}\right) = \frac{1}{\tilde{\kappa}\left(y_{(1)}^{(N)}\right)} \tilde{L}\left(c,a,\Omega,y_{(1)}^{(N)}\right) \qquad (6.83)$$

$$\tilde{\kappa}\left(y_{(1)}^{(N)}\right) = \int_{S_c}\int_{S_a}\int_{S_\Omega} \tilde{L}\left(c,a,\Omega,y_{(1)}^{(N)}\right) dc \cdot da \cdot d\Omega. \qquad (6.84)$$

The probability density (6.83) reflects our uncertainty about the unknown parameters c,a,Ω after the data $y_{(1)}^{(N)}$ have been observed.

If we are interested in the ML-estimates of the unknown parameters we have to find the point at which the likelihood function (6.82) reaches its maximum. This maximization can be decomposed into three steps:

(i) In the first step we shall find the maximum of the likelihood function over c for all possible values of a and Ω. The parameter c enters only the exponent in (6.82). It is easy to verify that this exponent can be rearranged in the following way.

$$\mathrm{tr}\left[\Omega\left(A\phi_{(N)}A^T - 2Am_{(N)}c^T + c\tau_{(N)}c^T\right)\right] =$$

$$= \mathrm{tr}\left\{\Omega\left[A\phi_{(N)}A^T + \left(c - \frac{1}{\tau_{(N)}}Am_{(N)}\right)\tau_{(N)}\left(c - \frac{1}{\tau_{(N)}}Am_{(N)}\right)^T - Am_{(N)}\frac{1}{\tau_{(N)}}m_{(N)}^T A^T\right]\right\}$$

$$= \mathrm{tr}[\Omega AH_{(N)}A] + \tau_{(N)}(c - Ab_{(N)})^T\Omega(c - Ab_{(N)}) \quad , \qquad (6.85)$$

where

$$b_{(N)} = \frac{1}{\tau_{(N)}} m_{(N)} \qquad (6.86)$$

$$b_{(N)i} = \frac{1}{\tau_{(N)}} m_{(N)i} = \frac{1}{t_{(N)} - t_{(1)}} \ell n \frac{f_{(N)i}}{f_{(1)i}} \qquad (6.87)$$

$$H_{(N)} = \phi_{(N)} - m_{(N)} \frac{1}{\tau_{(N)}} m_{(N)} = \phi_{(N)} - b_{(N)} \tau_{(N)} b_{(N)}^T$$

$$= \sum_{k=2}^{N} \frac{1}{T_{(k)}} (x_{(k)} - T_{(k)} b_{(N)}) (x_{(k)} - T_{(k)} b_{(N)})^T \tag{6.88}$$

$$H_{(N)ij} = \sum_{k=2}^{N} \frac{1}{T_{(k)}} \left(\ell n \frac{f_{(k)i}}{f_{(k-1)i}} - T_{(k)} b_{(N)i} \right) \left(\ell n \frac{f_{(k)j}}{f_{(k-1)j}} - T_{(k)} b_{(N)j} \right) . \tag{6.89}$$

As any possible Ω must be positive definite the last term in (6.85) can be only nonnegative. It is evident that the minimum of (6.85) and hereby the maximum of the likelihood function (6.82) is reached for

$$c = A b_{(N)} \tag{6.90}$$

$$c_{ri} = b_{(N)i} - a_{ri} b_{(N)r} \tag{6.91}$$

no matter what the values of the remaining parameters a, Ω are. After this first maximization step we have

$$\max_{c} \tilde{L}\left(c,a,\Omega,y_{(1)}^{(N)}\right) = \prod_{k=2}^{N} \left(f_{(k)}^T \tilde{a} \right) \cdot |\Omega|^{\frac{N-1}{2}} \exp \left\{ -\tfrac{1}{2} \operatorname{tr} \left[\Omega A H_{(N)} A^T \right] \right\} . \tag{6.92}$$

(ii) In the second step we shall find the matrix Ω which maximizes (6.71) for any a. According to Theorem 2 in Appendix A the maximum is taken on at

$$\Omega = (N-1) \cdot \left[A H_{(N)} A^T \right]^{-1} \tag{6.93}$$

and the maximum is

$$\max_{\Omega}\max_{c} L\left(c,a,\Omega,y_{(1)}^{(N)}\right) = e^{-(N-1)(n-1)/2}\frac{\prod_{k=2}^{N}\left(f_{(k)}^{T}\tilde{a}\right)}{\left|AH_{(N)}A^{T}\right|^{\frac{N-1}{2}}} \quad . \qquad (6.94)$$

(iii) In the third step we have to find the vector a (by which also A(6.58) and $\tilde{a}$ (6.57) are defined) which maximizes (6.94). This last step of maximization procedure cannot be solved analytically but only numerically. However, some rearrangement of the function (6.94) is suitable to facilitate the numerical solution. Theorem 3 in Appendix A makes it possible to rewrite (6.94) in the following form.

$$\max_{\Omega}\max_{c} \tilde{L}\left(c,a,\Omega,y_{(1)}^{(N)}\right) = \left(e^{n-1}|H|\right)^{-\frac{N-1}{2}} \cdot \lambda(\tilde{a}) \quad , \qquad (6.95)$$

where

$$\lambda(\tilde{a}) = \frac{\prod_{k=2}^{N} f_{(k)}^{T}\tilde{a}}{\left[\tilde{a}^{T}H_{(N)}^{-1}\tilde{a}\right]^{\frac{N-1}{2}}} \quad . \qquad (6.96)$$

Notice that the value of the function (6.96) remains unchanged when all components of the n-vector $\tilde{a}$ (6.57) are divided (or multiplied) by any positive number γ.

$$\bar{a} = \frac{1}{\gamma}\tilde{a} \qquad (6.97)$$

$$\lambda(\bar{a}) = \lambda(\tilde{a}) \quad .$$

For instance, if we choose $\gamma = \alpha_r$ then $\bar{a}_i = 1/\alpha_i$ for all i, including r. Alternatively, for $\gamma = \alpha_r/\bar{\alpha}$ we have $\bar{a}_i = a_i$ as defined by (5.30). This shows that *the ML-estimate of* $a_{ij} = \alpha_i/\alpha_j = a_{rj}/a_{ri}$ *is always the same no matter what technology is chosen as the reference competitor.* It also shows that it is impossible to estimate separately α_i, i = 1,2,...,n, but only their ratios.

A simple and effective numerical procedure which calculates the vector $\tilde{a}$ maximizing the function (6.96) is developed in Appendix B and the corresponding FORTRAN-subroutine AMLE can be found in Appendix C.

The whole numerical procedure for maximum likelihood estimation of all model parameters can be summarized into the following algorithm.

A l g o r i t h m MLEST:

1. Given the market shares $f_{(k)i}$, $i = 1,2,\ldots,n$; $k = 1,2,\ldots,N$, at time points t_k, $k = 1,2,\ldots,N$, calculate the scalar τ, the n-vector b and the symmetrical $n \times n$ - matrix H according to formulae (6.80), (6.87) and (6.89).

$$\tau = t_{(N)} - t_{(1)}$$

$$b_i = \frac{1}{\tau} \ln \frac{f_{(N)i}}{f_{(1)i}} \quad ; \qquad i = 1,2,\ldots,n$$

$$H_{ij} = \sum_{k=2}^{N} \frac{1}{T_{(k)}} \left(\ln \frac{f_{(k)i}}{f_{(k-1)i}} - T_{(k)} b_i \right) \left(\ln \frac{f_{(k)j}}{f_{(k-1)j}} - T_{(k)} b_j \right) ,$$

$$i,j = 1,2,\ldots,n \quad ,$$

where

$$T_{(k)} = t_{(k)} - t_{(k-1)} \quad .$$

2. Find the direction vector $\bar{a}$ of any length for which the function (6.96)

$$\frac{\prod_{k=2}^{N} f_{(k)}^{T} \bar{a}}{\left[\bar{a}^{T} H^{-1} \bar{a} \right]^{\frac{N-1}{2}}}$$

reaches its maximum. For this purpose the FORTRAN-subroutine AMLE, given in Appendix C, can be used. The subroutine AMLE gives the maximizing direction vector $\bar{a}$ of length (euclidian norm) equal to one.

3. For any chosen reference competitor (indexed by r) calculate the estimates a_{ir} and c_{ir}, i = 1,2,...,n, according to formulae

$$\hat{a}_{ir} = \frac{\bar{a}_r}{\bar{a}_i}$$

$$\hat{c}_{ir} = b_r - \hat{a}_{ir} b_i \quad . \tag{6.98}$$

4. Calculate the estimate of the covariance matrix $R = \Omega^{-1}$ according to formulae which follows from (6.93).

$$\hat{R}_{ij} = \frac{1}{N-1} \left(H_{ij} - \frac{H_{ir}}{\hat{a}_{jr}} - \frac{H_{rj}}{\hat{a}_{ir}} + \frac{H_{rr}}{\hat{a}_{ir}\hat{a}_{jr}} \right) \quad , \quad \forall i,j \neq r. \tag{6.99}$$

The FORTRAN - subroutine realizing this algorithm can be found in Appendix C.

Before examples of practical use of the algorithm MLEST are given, several additional comments are in order concerning the reliability of maximum likelihood estimates.

As stated above, the uncertainty of the various unknown parameters is fully characterized by the probability distribution (6.68) which -- in the case of a very flat prior distribution (6.71) -- is proportional to the likelihood function (6.81). The maximum likelihood estimate of the set of unknown parameters is the point at which this function reaches its maximum. This means that to a small region around this point a higher probability has been assigned by observed data than to the region of the same size around any other point in the space of all possible values of unknown parameters. It is evident that

the maximum likelihood estimates (as well as any other point estimates) can be very unreliable if the probability density function (6.81) is very flat or if it has a form of a ridge with almost the same height along some direction. We shall not go deeper into these details; however, we feel the necessity to emphasize that caution should be exercised in dealing with maximum likleihood estimates. The problem is not very critical in the case of the parameters c_{ir} but it may be very critical in the case of parameters a_{ir}.*) The reason can be seen well from the sensitivity analysis of the two-dimensional competition performed at the end of Section 5 (see Figures 12 and 13). This sensitivity analysis has shown that within the take-over time the model output is not very sensitive with respect to the ratio $a_{12} = \alpha_1/\alpha_2$. Conversely, the historical data contain little information about this parameter and when the data are noisy it may be very difficult or even impossible to extract this information. However, in this case the simplified model developed in Section 4 under the assumption that $a_{ir} = 1$ for all i may serve as a reasonable approximation.

For this reason the FORTRAN - subroutine MLEST listed in Appendix C provides the user both with the ML-estimate of all parameters (including a_{ir} if possible) and with ML-estimates of c_{ir} and R for $a_{ir} = 1$, $i = 1,2,\ldots,n$. The user has to choose the alternative which suits his case.

Example 3. Substitution of steam locomotives by diesel locomotives in the USA

Table 2, taken from Mansfield [9], gives the numbers of steam and Diesel locomotives in the USA in the years 1925 to 1959. The market shares within the time period 1939-1959 have been used as input data for parameter estimation. The result obtained by application of the subroutine MLEST is

$$\hat{a}_{12} = 1.56 \ , \quad \hat{c}_{12} = -0.505 \text{ year}^{-1} \ , \quad R = 0.75\ 10^{-2} \ .$$

*) In the next section it will be shown that, for the purpose of forecasting, there actually does not exist a single number (point estimate) by which the uncertain parameter a_{ir} could be replaced.

Table 2. Number of diesel and steam locomotives, USA, 1925-1959.

Source: E. Mansfield, Intrafirm Rates of Diffusion of an Innovation, Review of Economics and Statistics, 45 (1963), pp. 348-359.

YEAR (Dec.31)	Diesel Locomotives		Steam Locomotives	
	Number	Share	Number	Share
1925	1	*	67 713	**
1927	14	*	64 843	**
1929	25	*	60 572	**
1931	80	*	57 820	**
1933	85	*	53 302	**
1935	130	*	48 477	**
1937	293	*	46 342	**
1939	639	.0144	43 604	.9856
1941	1 517	.0349	41 911	.9651
1943	2 476	.0557	41 983	.9443
1945	4 301	.0949	41 018	.9051
1947	6 495	.1495	36 942	.8505
1949	12 025	.2838	30 344	.7162
1951	19 014	.4570	22 590	.5430
1953	24 209	.6636	12 274	.3364
1955	26 563	.8091	6 266	.1909
1957	29 137	.9179	2 608	.0821
1959	30 097	.9719	871	.0281

*Less than 1%
**More than 99%

To check the fit of the model with the true data the process has been recalculated using the subroutine PENETR with the parameter values modified according to (5.42) and (5.43)

$$\tilde{c}_1 = c_{12} \quad , \quad \tilde{c}_2 = \hat{c}_{22} = 0 \quad , \quad \tilde{\alpha}_1 = \hat{a}_{12} \quad , \quad \tilde{\alpha}_2 = a_{22} = 1, \; \rho = 0 \quad ,$$

and with the initial condition $f_{10} = 0.0144, f_{20} = 0.9856$, corresponding to the year 1939 (see Table 2). The calculated and the true process are compared in Figure 14a,b. The bended lines in Figure 14b indicate that the process is nonlogistic but, as it can be seen, within the take-over time (10% - 90%) the simple Fisher-Pry model could still be a reasonable approximation.

Example 4. Substitution of primary energy sources in world energy consumption.

In this example a typical sample of calculations performed on world energy data is recorded.

During the derivation of the model in Sections 4 and 5 it has been shown that the variations of the most important parameters c_{ir} (i = 1,2,...,n) are smoothed by integration so that the model output is not very sensitive to their fluctuations around some mean value. Nevertheless it is hard to believe that they could be even approximately constant for a very long period of time covering two world wars. To investigate this question several time periods have been considered for parameter estimation separately. In spite of the fact that the estimates obtained from different time periods are slightly different (estimates of a_{ij} sometimes very different), in all cases a remarkably good fit with all historical data, both forwards and backwards, has been obtained. This indicates a high stability of the process and supports the confidence in the forecast based on the model. To demonstrate this favorable feature of the model at least several of these calculations will be reported in detail so that the interested reader can easily reproduce them himself

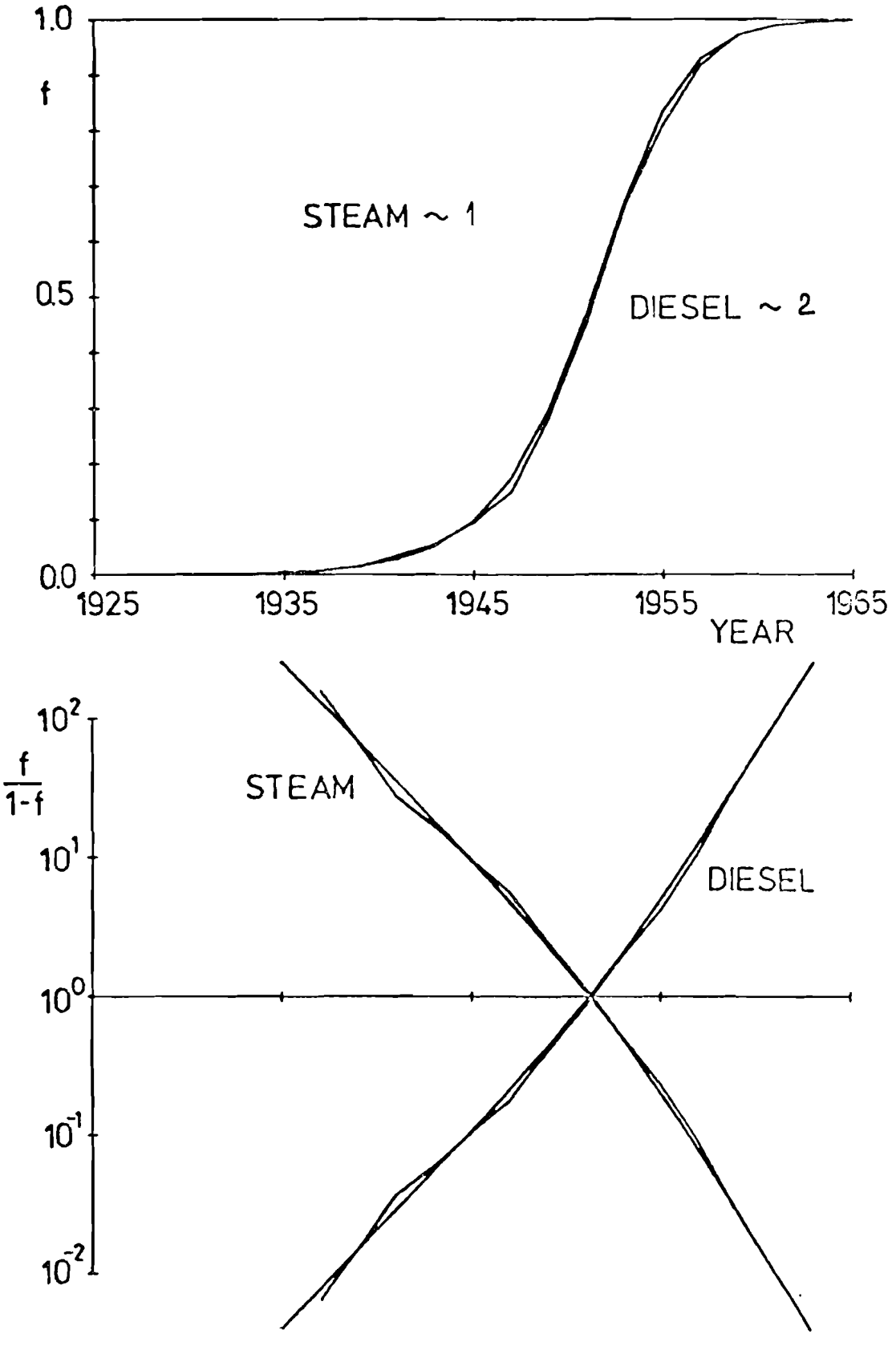


Figure 14a,b. Replacement of steam by diesel locomotives in the USA.

using the FORTRAN - subroutines in Appendix C.

In Table 3 the market shares of the main primary energy sources in the world energy consumption are registered for the time period 1920 to 1971. They have been calculated from world energy data collected at IIASA by N. Nakicenovic [12]. As the data on wood since 1951 have not been available they have been replaced by predictions obtained from the past history of the substitution process in order to make the table complete. Notice that these artificial data are relatively insignificant.

When all data from Table 3 are used the following estimates are obtained.

$$\hat{c}_{14} = 0.0884, \quad \hat{c}_{24} = 0.0601, \quad \hat{c}_{34} = 0.0353, \ (c_{44} = 0) \ ,$$

$$\hat{a}_{14} = 0.826, \quad \hat{a}_{24} = 0.867, \quad \hat{a}_{34} = 0.325, \ (a_{44} = 1) \ ,$$

where the indices have the following meaning: 1 ~ wood, 2 ~ coal, 3 ~ oil, 4 ~ natural gas, and one year is taken as time unit. Natural gas is chosen as the reference competitor; however, using the relations (5.13) and (5.15) the estimates can be easily recalculated to any other choice.

The smoothed curves in Figures 15a,b show the output of the deterministic model (subroutine PENETR with parameter values $\tilde{c}_i = \hat{c}_{i4}$, $\tilde{\alpha}_i = \hat{a}_{i4}$, $\rho = 0$) for initial market shares in 1920 taken from Table 3.

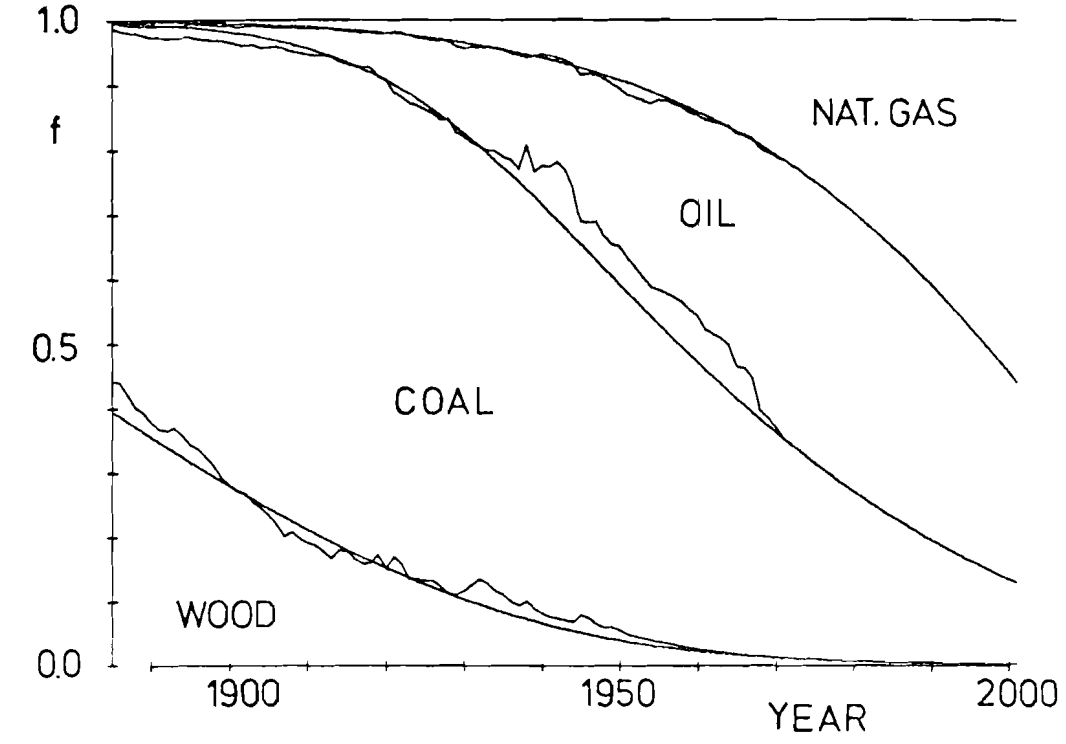
As discussed above, the estimates of the parameters a_{ir} are rather unreliable but also not very significant from the point of view of the model output. If all of them are set to one the following estimates of the parameters c_{ir} are obtained (subroutine MLEST in Appendix C gives both these results at the same time):

$$\hat{c}_{14} = 0.0973, \quad \hat{c}_{24} = 0.0622, \quad \hat{c}_{34} = 0.0119, \ (c_{44} = 0) \ .$$

Table 3. Market shares of primary energy sources in world energy consumption.

Year	Wood	Coal	Oil	Nat.Gas
1920	0.15118	0.75531	0.07347	0.02004
1921	0.16953	0.71997	0.09197	0.01853
1922	0.15714	0.72469	0.09805	0.02013
1923	0.13493	0.73715	0.10397	0.02395
1924	0.13537	0.73278	0.10477	0.02707
1925	0.13283	0.72967	0.10942	0.02809
1926	0.13216	0.72456	0.11228	0.03099
1927	0.12342	0.72286	0.12176	0.03196
1928	0.11317	0.73508	0.11934	0.03241
1929	0.11043	0.71677	0.13292	0.03988
1930	0.11648	0.70440	0.13571	0.04341
1931	0.12659	0.68565	0.14597	0.04179
1932	0.13488	0.66977	0.15282	0.04252
1933	0.12825	0.66984	0.16063	0.04127
1934	0.11688	0.68182	0.15762	0.04368
1935	0.11048	0.67938	0.16515	0.04499
1936	0.10005	0.68465	0.16448	0.05081
1937	0.09404	0.67701	0.17578	0.05318
1938	0.10071	0.70662	0.13574	0.05692
1939	0.08825	0.67869	0.17845	0.05461
1940	0.08044	0.69424	0.17230	0.05302
1941	0.07611	0.69820	0.17114	0.05455
1942	0.07422	0.70751	0.16118	0.05709
1943	0.07073	0.69876	0.16873	0.06178
1944	0.06853	0.67301	0.19205	0.06641
1945	0.07918	0.61006	0.22659	0.08416
1946	0.07444	0.61404	0.22807	0.08345
1947	0.06510	0.62398	0.22698	0.08394
1948	0.06028	0.60720	0.24353	0.08900
1949	0.06115	0.59318	0.24834	0.09734
1950	0.05515	0.59671	0.24358	0.10456
1951	0.04902 *)	0.58497	0.25224	0.11377
1952	0.04588	0.57268	0.26235	0.11909
1953	0.04314	0.56042	0.27285	0.12359
1954	0.04062	0.54817	0.28298	0.12823
1955	0.03801	0.54635	0.29295	0.12268
1956	0.03542	0.54239	0.29791	0.12427
1957	0.03321	0.53893	0.29735	0.12950
1958	0.03097	0.53441	0.30095	0.13367
1959	0.02899	0.52412	0.30542	0.14146
1960	0.02689	0.51580	0.31156	0.14575
1961	0.02493	0.49792	0.32364	0.15351
1962	0.02331	0.49191	0.32743	0.15735
1963	0.02150	0.48724	0.33193	0.15933
1964	0.01994	0.47731	0.33871	0.16404
1965	0.01861	0.44632	0.36116	0.17390
1966	0.01709	0.44499	0.36191	0.17611
1967	0.01591	0.43108	0.37146	0.18154
1968	0.01459	0.42246	0.37647	0.18648
1969	0.01353	0.37173	0.41092	0.20381
1970	0.01229	0.35786	0.42092	0.20894
1971	0.01141	0.34056	0.43216	0.21587

*The data on wood consumption for energy production since 1951 have not been available. These numbers are estimates based on the past history of the substitution process (1885-1950).



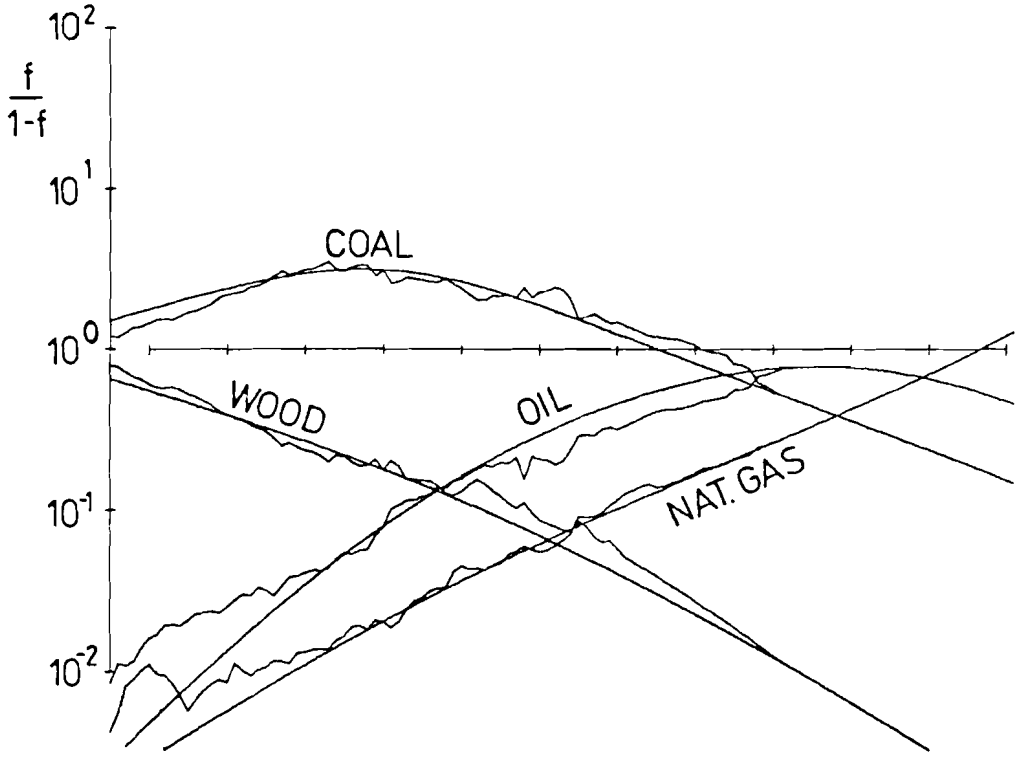


Figure 15a.b. Primary energy substitution: all model parameters estimated from the period 1920-1971.

The corresponding model output for the starting point in 1920 is shown in Figures 16a,b. In spite of the great difference in parameters a_{i4} (especially in a_{34}) the difference in the model outputs is not so great as one could expect. The comparison of fits with historical data indicates that in this case the assumption $a_{i4} = 1$ (for all i) may be more realistic than to rely on uncertain ML-estimates of these parameters. The important question, how the uncertainty of parameters has to be projected into the future, will be studied in detail in the next section where we shall deal with forecasting.

One could expect that the 2nd World War and the intensive technical development after this war could cause a significant change in the model parameters. If this were true then the prewar data should not be used in forecasting the future development of the substitution process. To clear up this question only the postwar data from the time period 1945 to 1971 have been considered in parameter estimation and the model has been used to "backcast" the past history of the process. Due to the relatively low number of data with narrow range of their amplitudes only the simplified model under the assumption $a_{ir} = 1$ (for all i) could be obtained in this case. The estimates of the remaining free parameters are

$$\hat{c}_{14} = 0.1107, \quad \hat{c}_{24} = 0.0586, \quad \hat{c}_{34} = 0.0114, \quad (c_{44} = 0) \ .$$

The comparison with historical data, for the starting point in 1971, is given in Figures 17a,b. The bad fit of the curve separating the market shares of wood and coal in Figure 17a cannot be considered as a failure. Notice that only six points of used data on wood (1945 to 1950) are true. The remaining 21 points are artificial and smoothed and were taken by the estimation procedure as very precise. This, of course, drastically influenced the result of estimation. To overcome this difficulty, wood (which is insignificant for the future development of the process) and coal have been aggregated (by summing up the two corresponding columns in Table 3) and considered as one competitor.

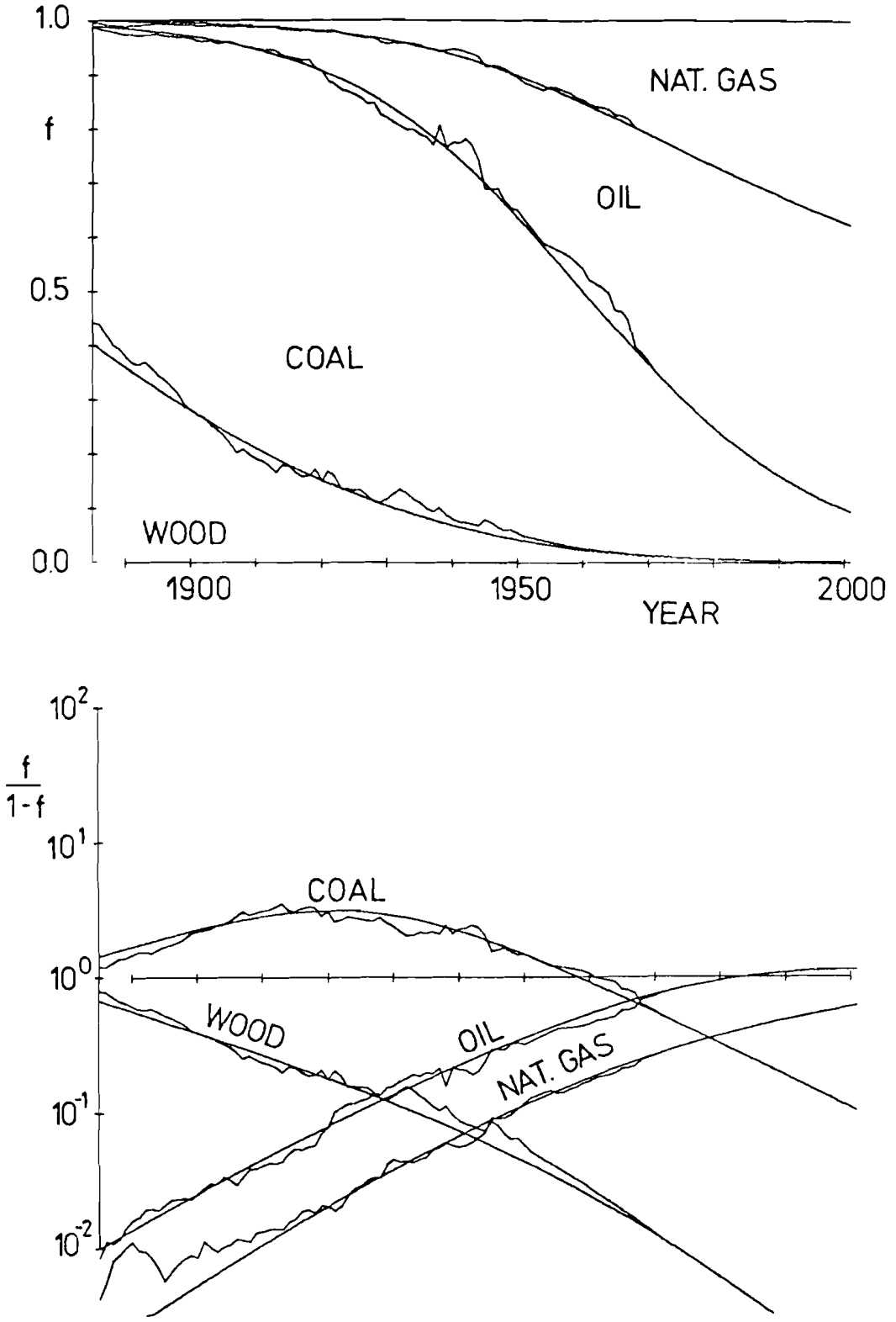


Figure 16a,b. Primary energy substituion: parameters a set to 1; remaining estimated from the period 1920-1971.

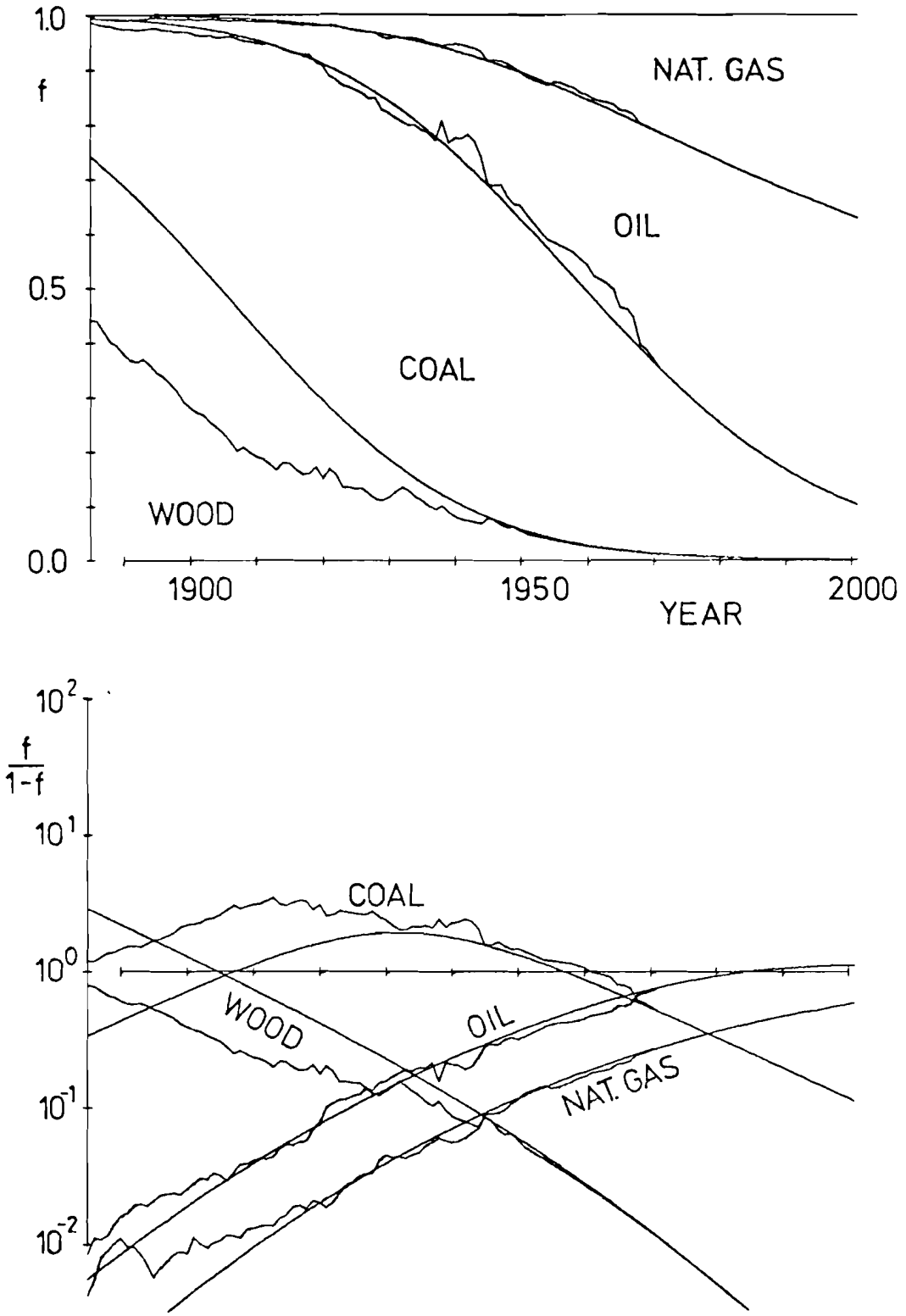


Figure 17a,b. Primary energy substitution: parameters a set to 1; remaining estimated from postwar data 1945-1971.

The result of the repeated calculation is

$$\hat{c}_{13} = 0.0621 \quad , \quad \hat{c}_{23} = 0.0114 \quad , \quad (c_{33} = 0) \quad ,$$

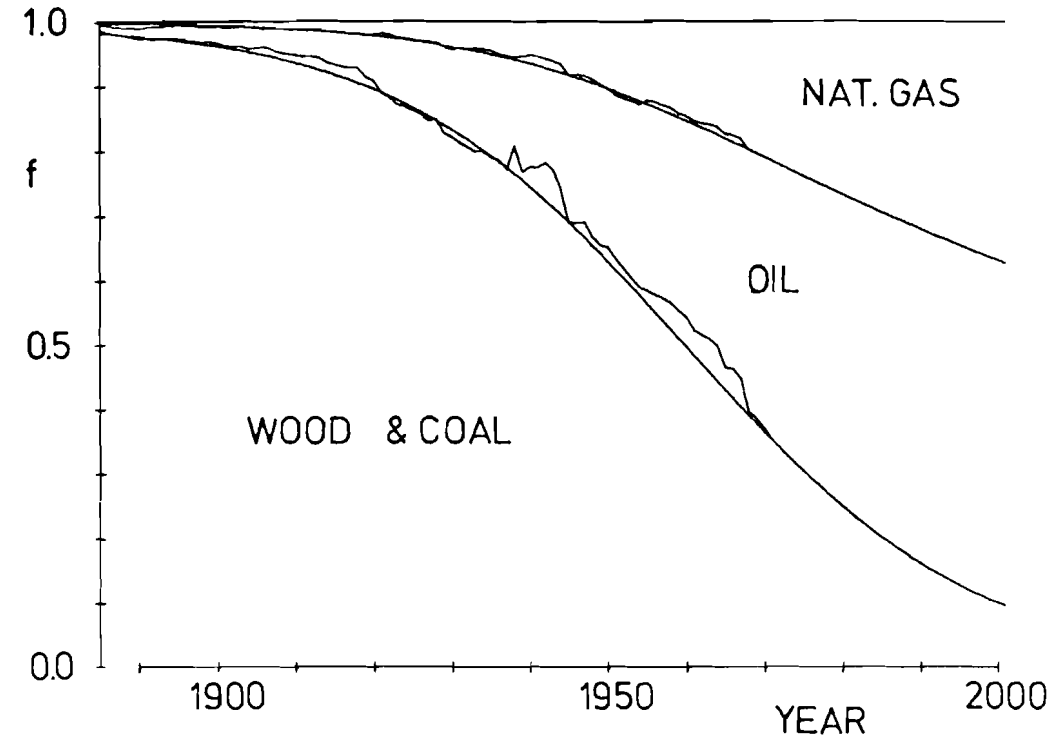
where the indices mean: 1 ~ wood + coal, 2 ~ oil, 3 ~ natural gas. Considering the "goodness-of-fit" with historical data for 70 years backwards, shown in Figures 18a,b, it is hard to believe that the law governing the substitution process should be much different in the near future. The opponent, who is of different opinion, should attack the assumptions on which the model is based and show why and how much, whether significantly, they will be violated. Of course, the future development of the substitution process can be considerably influenced when a new and competitive energy source, like nuclear, enters the market. This will be shown in the next section.

7. FORECASTING

In this section the problem of forecasting is studied with emphasis on the following objectives:

- to clear up the relation between the deterministic and probabilistic models developed in previous sections in order to give a precise probabilistic meaning to the curves generated by the deterministic model, which -- of course -- can never be precisely true;
- to investigate the suitability of maximum likelihood estimates for the purpose of forecasting;
- to show how the uncertainty of parameters can be projected into the future;
- to show, by the example of nuclear energy, how a new technology can be incorporated into the model.

To follow this program let us show first that the output of the deterministic model, calculated for some particular time point t and for given parameters and initial conditions, is



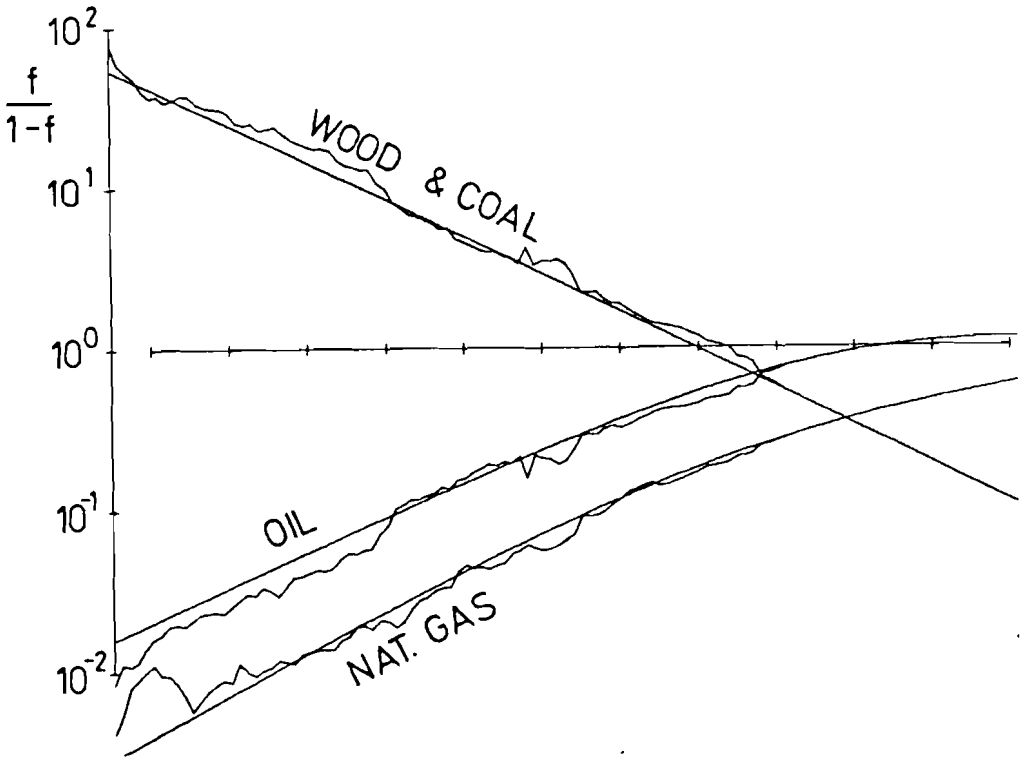


Figure 18a.b. Wood and coal aggregated, parameters a set to 1; remaining estimated from postwar data 1945-1971.

neither the maximum of the probability density nor the mean value but the median. In the two-dimensional case (n = 2) it means that with probability one half, the possible true value lies on one side of the calculated curve and with the same probability on the other side. Before we prove this statement, let us formulate it more precisely and generally for the multivariate case.

The output of the process at time t_{k+1} is fully determined by the vector $y_{(k+1)}$ ($y_{(k+1)i} = f(t_{k+1})_i, i \neq r$). Let us denote the output of the deterministic model by $y^*_{(k+1)}$ to be distinguished from any other possible output $y_{(k+1)}$. Further, let us introduce the following spaces of possible process outputs.

$$y^-_{(k+1)} = \{ y_{(k+1)} : y_{(k+1)i} \leq y^*_{(k+1)i}, \forall i \neq r \}$$

$$y^+_{(k+1)} = \{ y_{(k+1)} : y_{(k+1)i} > y^*_{(k+1)i}, \forall i \neq r \} .$$

What we are stating is: *Given the model parameters* $K = \{c,a,\Omega\}$ *and the initial condition* $y_{(k)}$, *the probability that the output* $y_{(k+1)}$ *will lie in the space* $y^-_{(k+1)}$ *is the same as the probability that the output will lie in the space* $y^+_{(k+1)}$.

$$Pr(y_{(k+1)} \in y^-_{(k+1)} | y_{(k)}, K) = Pr(y_{(k+1)} \in y^+_{(k+1)} | y_{(k)}, K) \quad . \qquad (7.1)$$

To prove this statement let us calculate the probability on the left hand side of (7.1). Using the regular transformation between the random vectors $y_{(k+1)}$ and $e_{(k+1)}$, and considering that $y^*_{(k+1)}$ is calculated for $e_{(k+1)} = 0$, we obtain

$$\Pr\left(y_{(k+1)} \in \bar{y}_{(k+1)} \mid y_{(k)}, K\right) = \int_{0}^{y^*_{(k+1)}} p\left(y_{(k+1)} \mid y_{(k)}, K\right) dy_{(k+1)}$$

$$= \int_{-\infty}^{0} p\left(e_{(k+1)} \mid y_{(k)}, K\right) de_{(k+1)} = \int_{0}^{\infty} p\left(e_{(k+1)} \mid y_{(k)}, K\right) de_{(k+1)} \cdot \quad (7.2)$$

The last equality in (7.2) follows from the fact that the distribution of $e_{(k+1)}$ is normal, according to assumption (6.53), and proves the statement.

Now, let us follow the question of what happens when some or all model parameters are not precisely known; how is the uncertainty of the parameters reflected in the uncertainty of the forecast. The precise Bayesian answer to this question is as follows.

Let $y_{(N)}$ be the last output of the process which is known and $y_{(N+1)}$ be the future output at time $t_{N+1} > t_N$ we want to forecast. If the set of all model parameters K were known, then all that could be said about the future output is contained in the probability density $p\left(y_{(N+1)} \mid y_{(N)}, K\right)$ defined by our model. When all or some of the parameters are unknown or uncertain this probability density is not available. Consider the parameter set K decomposed into two subsets

$$K = \{K_u, K_c\} \quad , \quad (7.3)$$

where K_u is the subset of unknown parameters while the parameters K_c are considered as certain. Not knowing the parameters K_u we have to look for the probability density $p\left(y_{(N+1)} \mid y_{(1)}^{(N)}, K_c\right)$ where the information about the unknown parameters is replaced by the information contained in the known past history of the process. Forecasting under the lack of parameter values is nothing else than calculation of this probability density which can be performed using two basic formulae (6.2) and (6.1) in the following way.

$$p\left(y_{(N+1)} \mid y_{(1)}^{(N)}, K_c\right) = \int p\left(y_{(N+1)} \mid y_{(1)}^{(N)}, K_c, K_u\right) p\left(K_u \mid y_{(1)}^{(N)}, K_c\right) dK_u \quad . \tag{7.4}$$

In our case, both probability density functions which are required on the right-hand side of (7.4), are already known. The first one is given by (6.66), the second one -- for all parameters unknown -- by (6.83), (6.82) and (6.84).

The formula (7.4) indicates under what conditions the unknown parameters K can be simply replaced by their maximum likelihood estimates. Consider the probability density $p(y_{(N+1)} \mid y_{(N)}, K_c, K_u)$ as a function of K_u for given $y_{(N)}$ and any but fixed $y_{(N+1)}$. If the probability distribution for K_u is highly concentrated around the ML-estimate $\hat{K}_u$, as shown in Figure 19a, it is evident that a good approximation of the integral (7.4) can be obtained if the variable K_u in the first part of the integrand in (7.4) is simply replaced by the fixed point $\hat{K}_u$.

$$p\left(y_{(N+1)} \mid y_{(1)}^{(N)}, K_c\right) \approx p(y_{(N+1)} \mid y_{(N)}, K_c, K_u)\Big|_{K_u = \hat{K}_u} \quad . \tag{7.5}$$

However, if the situation is like Figure 19b, the approximation (7.5) does not hold and the integration in (7.4) has to be performed. Unfortunately, it is not easy to recognize what situation occurs without a more detailed investigation.

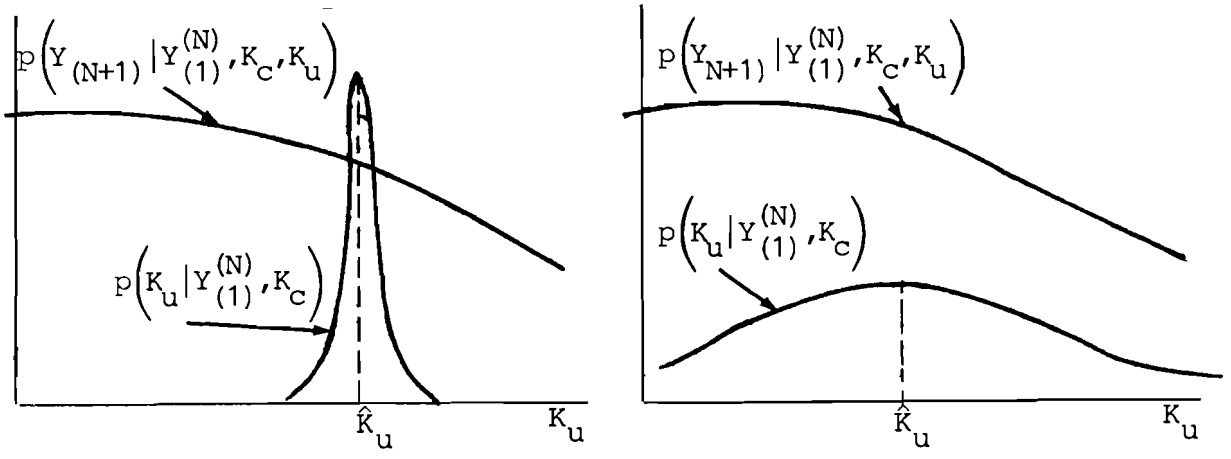


Figure 19a,b. Two extreme situations in Bayesian forecasting.

In our case the integration (7.4) can be performed analytically for the unknown parameters $K_u = \{c,\Omega\}$ but only numerically when also the parameter vector a is considered as unknown. To facilitate the integration let us bring the formula (7.4) into a more convenient form. When the density $p(K_u|y_{(1)}^{(N)},K_c)$ is expressed by the Bayes formula (analogous to (6.70))

$$p\left(K_u|y_{(1)}^{(N)},K_c\right) = \frac{p\left(y_{(2)}^{(N)}|y_{(1)},K_c,K_u\right)p\,(K_u|K_c,y_{(1)})}{\int p\left(y_{(2)}^{(N)}|y_{(1)},K_c,K_u\right)p(K_u|K_c,y_{(1)})\,dK_u} \tag{7.6}$$

and the relation (again the basic formula (6.1))

$$p\left(y_{(N+1)}|y_{(1)}^{(N)},K_c,K_u\right)p\left(y_{(2)}^{(N)}|y_{(1)},K_c,K_u\right) = p\left(y_{(2)}^{(N+1)}|y_{(1)},K_c,K_u\right) \tag{7.7}$$

is applied, the formula (7.4) gets the form

$$p\left(y_{(N+1)}|y_{(1)}^{(N)},K_c\right) = \frac{\int p\left(y_{(2)}^{(N+1)}|y_{(1)},K_c,K_u\right)p\,(K_u|K_c,y_{(1)})dK_u}{\int p\left(y_{(2)}^{(N)}|y_{(1)},K_c,K_u\right)p(K_u|K_c,y_{(1)})dK_u}\,. \tag{7.8}$$

Notice that the initial condition $y_{(1)}$ for our stochastic model can be, actually, considered as one of its parameters which is known. If it can be assumed that the prior distribution for the unknown parameters $p(K_u|y_{(1)},K_c)$ is very flat even when K_c and $y_{(1)}$ are a priori known, then -- under conditions specified in Section 6 -- the following result is obtained.

$$p\left(y_{(N+1)}|y_{(1)}^{(N)},K_c\right) = \frac{\int p\left(y_{(2)}^{(N+1)}|y_{(1)},K_c,K_u\right)dK_u}{\int p\left(y_{(2)}^{(N)}|y_{(1)},K_c,K_u\right)dK_u}\,. \tag{7.9}$$

Notice that the probability densities in (7.9) are, in fact, likelihood functions given by formula (6.81).

$$p\left(y_{(2)}^{(N+1)} \mid y_{(1)}, K_c, K_u\right) = L\left(c, a, \Omega, y_{(1)}^{(N+1)}\right) \quad , \tag{7.10}$$

$$p\left(y_{(2)}^{(N)} \mid y_{(1)}, K_c, K_u\right) = L\left(c, a, \Omega, y_{(1)}^{(N)}\right) \quad . \tag{7.11}$$

Of course, the variable $y_{(N+1)}$ incorporated into (7.10) has to be considered not as fixed but free.

We shall exploit the general formula (7.9) in two steps. First, it will be assumed that the parameter vector a is known. In the second step the solution will be generalized for the case when also a is unknown.

7.1 Parameter Vector a Known

In this case we have

$$K_u = \{c, \Omega\} \quad , \quad K_c = \{a\}$$

and the formula (7.9) can be written

$$p\left(y_{(N+1)} \mid y_{(1)}^{(N)}, a\right) = \frac{\kappa\left(y_{(1)}^{(N+1)}, a\right)}{\kappa\left(y_{(1)}^{(N)}, a\right)} \tag{7.12}$$

$$\kappa\left(y_{(1)}^{(M)}, a\right) = \int_{S_c} \int_{S_\Omega} L\left(c, a, \Omega, y_{(1)}^{(M)}\right) dc \cdot d\Omega \quad , \tag{7.13}$$

where M stays either for N + 1 or N.

The ratio on the right-hand side of (7.12) is given by Theorem 5 in Appendix A. According to this theorem the probability distribution for the future output of the process, given only the past history of the process and parameter vector a, is (for the sake of simplicity we omit the normalizing factor in (A.40)

$$p\left(y_{(N+1)} \mid y_{(1)}^{(N)}, a\right) \propto \frac{f_{(N+1)}^{T}\,\tilde{a}}{\Pi\left(f_{(N+1)}\right)} \frac{1}{\left(1 + \hat{e}_{(N+1)}^{T}\,\Psi_{(N)}^{-1}\,\hat{e}_{(N+1)}\right)^{\frac{N+n}{2}}} , \quad (7.14)$$

where

$$\hat{e}_{(N+1)} = A\,x_{(N+1)} - \hat{c}_{(N)}\,(t_{N+1} - t_N) \quad (7.15)$$

$$x_{(N+1)i} = \ell n\,\frac{f_{(N+1)i}}{f_{(N)i}} , \qquad \forall i \quad (7.16)$$

$$\Psi_{(N)} = (t_{N+1} - t_N)\,\frac{t_{N+1} - t_1}{t_N - t_1}\,A\,H_{(N)}\,A^{T} \quad (7.17)$$

and $\hat{c}_{(N)}$ is the maximum likelihood estimate (6.98) of the parameter vector c available at time t_N.

Notice that the transformation between the random variable $y_{(N+1)}$ and the variable $\hat{e}_{(N+1)}$ introduced by (7.15) is the same as in (6.65); only the true parameter vector c is replaced by its ML-estimate. Using this transformation, the Jacobian of which is reciprocal to (6.64), we can calculate

$$p\left(\hat{e}_{(N+1)} \mid y_{(1)}^{(N)}, a\right) \propto \frac{1}{\left(1 + \hat{e}_{(N+1)}^{T}\,\Psi_{(N)}^{-1}\,\hat{e}_{(N+1)}\right)^{\frac{N+n}{2}}} . \quad (7.18)$$

Hence, the random variable $\hat{e}_{(N+1)}$ has the Student's t-distribution with zero mean and with the covariance matrix

$$E\left[\hat{e}_{(N+1)}\hat{e}_{(N+1)}^{T} \mid y_{(1)}^{(N)}, a\right] = \frac{1}{N-2}\,\Psi_{(N)} = (t_{N+1} - t_N)\,\hat{\hat{R}}_{(N)} \quad (7.19)$$

where

$$\hat{\hat{R}}_{(N)} = \frac{1}{N-1}\,\frac{t_{N+1} - t_1}{t_N - t_1}\,A\,H_{(N)}\,A^{T} \quad (7.20)$$

$$\hat{\hat{R}}_{(N)ij} = \frac{1}{N-1}\frac{t_{N+1}-t_1}{t_N-t_1}\left(H_{ij} - \frac{H_{ir}}{a_{jr}} - \frac{H_{rj}}{a_{ir}} + \frac{H_{rr}}{a_{ir}a_{jr}}\right) \quad . \qquad (7.21)$$

Notice that $\hat{\hat{R}}_{(N)}$ is related to the ML-estimate $\hat{R}_{(N)} = \hat{\Omega}^{-1}_{(N)}$ (6.93) in the following way:

$$\hat{\hat{R}}_{(N)} = \frac{t_{N+1}-t_1}{t_N-t_1}\hat{R}_{(N)} \quad . \qquad (7.22)$$

From these theoretical results the following practical conclusions can be drawn:

- *The deterministic model in which the unknown parameters* c_{ir}, $\forall i \neq r$, *are replaced by their* ML-*estimates, gives the point of the distribution (7.14) where* $\hat{e}_{(N+1)} = 0$. *As the Student's* t-*distribution (7.18) is symmetric (and for* $N >> n-1$ *close to normal) the deterministic forecast has the same meaning as if these parameters were known, i.e. the meaning of median.*

- *The increase of uncertainty in the forecast due to uncertain parameters* c *and* Ω *is reflected in the covariance matrix (7.22) and in the change of the shape of the probability distribution* $p(e_{(N+1)}|Y_{(N)},K)$ *from the normal to Student's.*

- *For given parameter vector* a *the best estimate of* c, *for the purpose of forecasting, is the* ML-*estimate. The* ML-*estimate of the covariance matrix* R *has to be corrected according to (7.22).*

Unfortunately, this is also not true when the parameter vector a is unknown, as we shall show now.

7.2 Parameter Vector a Unknown

If no parameters can be considered as known, then the parameter sets in the general formula (7.9) are

$$K_u = \{c,a,\Omega\} \quad , \qquad K_c = \emptyset \quad .$$

and, according to the general formula (7.9), instead of (7.12) we now have

$$p\left(y_{(N+1)} \; y^{(N)}_{(1)}\right) = \frac{\int_{S_a} \kappa\left(y^{(N+1)}_{(1)}, a\right) da}{\int_{S_a} \kappa\left(y^{(N)}_{(1)}, a\right) da} \quad . \tag{7.23}$$

Using the formulae (A.37) from Theorem 5 and (A.16) from Theorem 3 (Appendix A), it can easily be verified that the following proportionality holds:

$$p\left(y_{(N+1)} \,|\, y^{(N)}_{(1)}\right) \propto \frac{1}{\Pi(f_{(N+1)})} \int_{S_a} \frac{\prod_{k=2}^{N+1} (f^T_{(k)} \tilde{a})}{\left(|H_{(N+1)}| \cdot \tilde{a}^T H^{-1}_{(N+1)} \tilde{a}\right)^{\frac{N+n}{2}}} da. \tag{7.24}$$

For any chosen $y_{(N+1)}$ (or $f_{(N+1)}$ fulfilling the condition $\sum f_{(N+1)i} = 1$) the right hand side of (7.14) can be evaluated by numerical integration. In this way the entire probability density (as a function of the variable vector $y_{(N+1)}$) can be obtained in the form of a numerical table and any of its characteristics can be calculated numerically. This numerical calculation can be facilitated using the formulae proved in Theorem 4 (Appendix A):

$$H^{-1}_{(N+1)} = H^{-1}_{(N)} - \frac{1}{\gamma_{(N+1)}} H^{-1}_{(N)} d_{(N+1)} d^T_{(N+1)} H^{-1}_{(N)} \tag{7.25}$$

$$d_{(N+1)} = x_{(N+1)} - b_{(N)} (t_{N+1} - t_N) \tag{7.26}$$

$$\gamma_{(N+1)} = \theta_{(N+1)} + d^T_{(N+1)} H^{-1}_{(N)} d_{(N+1)} \tag{7.27}$$

$$\theta_{(N+1)} = (t_{N+1} - t_N)(t_{N+1} - t_1)/(t_N - t_1) \tag{7.28}$$

$$|H_{(N+1)}| = \theta_{(N+1)} |H_{(N)}| \gamma_{(N+1)} \quad , \tag{7.29}$$

where $x_{(N+1)}$ is defined by (7.16). Notice, that the determinant (7.29) can be taken out from the integral (7.24) and $\theta_{(N+1)} |H_{(N)}|$ can be omitted being a part of the normalizing factor.

The following practical conclusions have to be drawn:

- *Generally, there does not exist a single number - a point estimate - by which the unknown parameter* a *could be replaced for the purpose of forecasting (neither in deterministic nor in probabilistic models). An exception is the situation shown in Figure 19a, which occurred in Example 3 (Diesel versus steam locomotives in the USA). In other cases the maximum likelihood estimate of the parameter vector* a *has to be handled with caution. Fortunately, within the most important period of transition the model output is not very sensitive with respect to this parameter, as it was shown by the sensitivity analysis of the two-dimensional case in Section 5. Usually, the assumption* $a_{ir} = 1$, $\forall i$, *can serve as a good approximation. Because of these reasons the subroutine MLEST in Appendix C provides the user both with ML-estimates of all parameters and with ML-estimates of the parameter vector* c *and covariance matrix* R *under the assumption* $a_{ir} = 1$, $\forall i$. *Whether the question is critical, it can be recognized by plotting the ratios of market shares versus time in semilogarithmic scale, like in Figures 7a,b,c,d.*

- *If the user has a reason to assume some other value of the parameter* a *than suggested, he can recalculate the estimates* $\hat{c}$ *and* $\hat{R}$ *using the formulae (6.98) and (7.21). The vector* b *and the matrix* H, *which enter these formulae are also supplied by the subroutine MLEST.*

Now, following these rules, we shall show by the example of nuclear energy, how a new technology can be incorporated into the model.

Example 5. Forecast of market penetration by nuclear energy.

In Example 4 only four main competitors in the world energy market have been considered: wood (i = 1), coal (i = 2), oil (i = 3) and natural gas (i = 4). If we want to forecast the future development of this substitution process we have to consider the

possible newcomers. In this example it will be shown how this can be done in the case of nuclear energy (i = 5). However, the same or similar procedure can be applied for any other additional competitive energy source.

To be able to apply the deterministic model (with the probabilistic interpretation given above) we have to determine the parameters c_{ij}, a_{ij} characterizing the relation between each pair of competitors. Only 2(n-1) = 8 of them are independent, the rest being determined by relations (5.13) to (5.15). If, for instance, natural gas is considered as the reference competitor (r = 4) then the full set of model parameters is

$$c_{i4}, a_{i4} : i = 1,2,3,5 \quad .$$

If we have no reasons to expect a significant change in parameters characterizing the competition between the main existing energy sources -- as shown in Example 4 they could be considered as constants for more than the last seventy years -- we can use their estimates based on historical data. Using the data from the time period 1920 to 1971 we have (see Example 4) for $a_{14} = a_{24} = a_{34} = 1$

$$\hat{c}_{14} = 0.0973 \quad , \quad \hat{c}_{24} = 0.0622 \quad , \quad \hat{c}_{34} = 0.0119 \quad , \quad (c_{44} = 0).$$

Notice that wood is no longer significant; it is considered only for completeness.

To complete the set of parameters we have to determine the remaining parameters c_{54} and a_{54}. Having almost no historical experience with nuclear energy we have to use the economic assessment of this newcomer relative to some existing and significant competitor. For the purpose of demonstration the comparison between nuclear energy and oil, as primary energy sources for electricity production, given by FRG Ministry for Research and Development [19] has been used. According to this source the specific investments and total production costs for LWR-nuclear plant and natural gas plant, both for a base load

of 4000 h/a, are:

nuclear - α_5 = 1150 DM/kW, c_5 = 0.063 DM/kWh = 552 DM/kW a

gas - α_4 = 755 DM/kW, c_4 = 0.075 DM/kWh = 657 DM/kW a .

Expecting that the competition will take place -- at least in the first stage -- in the field of electricity production, the growth rate factor $\rho = 0.06\ a^{-1}$ has been assumed. Substitution of these figures into the formulae (5.11) and (5.12) gives

$$a_{54} = \frac{\alpha_5}{\alpha_4} = 1.52$$

$$c_{54} = \frac{c_5 - c_4}{\alpha_4} + \left(\frac{\alpha_5}{\alpha_4} - 1\right)\rho = -0.139 + 0.031 = -0.108\ a^{-1} .$$

Taking the last known market shares in the year 1971 as initial conditions (f_{10} = 0.0114, f_{20} = 0.3406, f_{30} = 0.4322, f_{40} = 0.2159, f_{50} = 0) the market shares in the time period 1885-1973 have been calculated using the subroutine PENETR with parameter values modified according to (5.42) and (5.43). In the year 1973 1% market share of nuclear energy, (f_5 = 0.01) has been introduced and the remaining market shares have been corrected accordingly. For these new initial conditions the substitution process has been calculated until the year 2050. The result is plotted in Figure 20a,b.

The economical assessments performed by various authors diverge considerably. To see how much the forecast is affected when rather different input data are used, the calculation has been repeated using the assessment given by Michaelis [20] for electricity power plants using LWR - nuclear energy and oil as primary energy sources. According to this source the specific investments and total production costs for the base load 7000 h/a are:

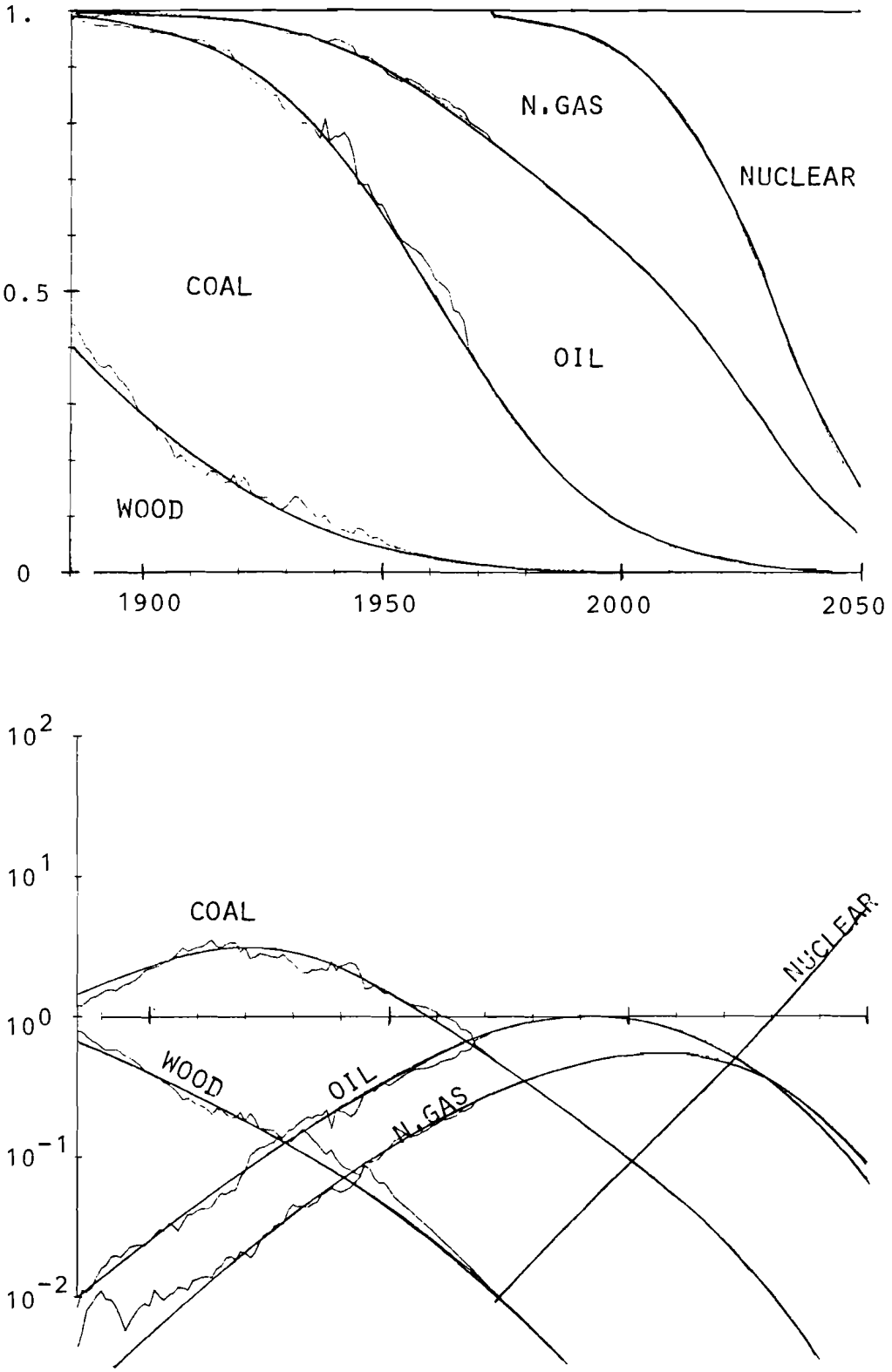


Figure 20. Forecast of market penetration by nuclear energy--economic assessment from [19].

nuclear - α_5 = 1500 DM/kW, c_5 = 0.0429 DM/kWh = 376 DM/kW a

oil - α_3 = 720 DM/kW, c_3 = 0.0639 DM/kWh = 560 DM/kW a .

For these figures and $\rho = 0.06\ a^{-1}$ the formulae (5.11) and (5.12) give

$$a_{53} = \frac{\alpha_5}{\alpha_3} = 2.08$$

$$c_{53} = \frac{c_5 - c_3}{\alpha_3} + \left(\frac{\alpha_5}{\alpha_3} - 1\right)\rho = -0.256 + 0.065 = -0.191\ a^{-1} .$$

Choosing oil as the reference competitor in this case, the estimates of the parameters c_{ir}, given above for r = 4, have been recalculated for r = 3 (oil) using the formula $\hat{c}_{i3} = \hat{a}_{43}(\hat{c}_{i4} - \hat{c}_{34})$ following from (5.15).

$$\hat{c}_{13} = 0.0854, \quad \hat{c}_{23} = 0.0504, \quad (c_{33} = 0), \quad \hat{c}_{43} = -0.0119 .$$

The corresponding model output, obtained for the same initial conditions as in the previous case, is plotted in Figures 21a,b. To make the comparison complete the parameter values

$$\hat{a}_{54} = 2.08 , \qquad \hat{c}_{54} = -0.172$$

have also been determined using the formulae (5.11) and (5.12). Notice the great difference in these parameter values when compared with the previous case. In spite of this great difference the model outputs given in Figures 20a,b and Figures 21a,b do not differ so drastically as one could expect. This again shows the very high stability of the substitution process and explains the technological "fate" observed by Marchetti [2].

The possible objection that the economical assessment made for FRG may not be representative for the whole world is fully justified. Nevertheless, the rather low sensitivity of the model output with respect to its parameter values demonstrated

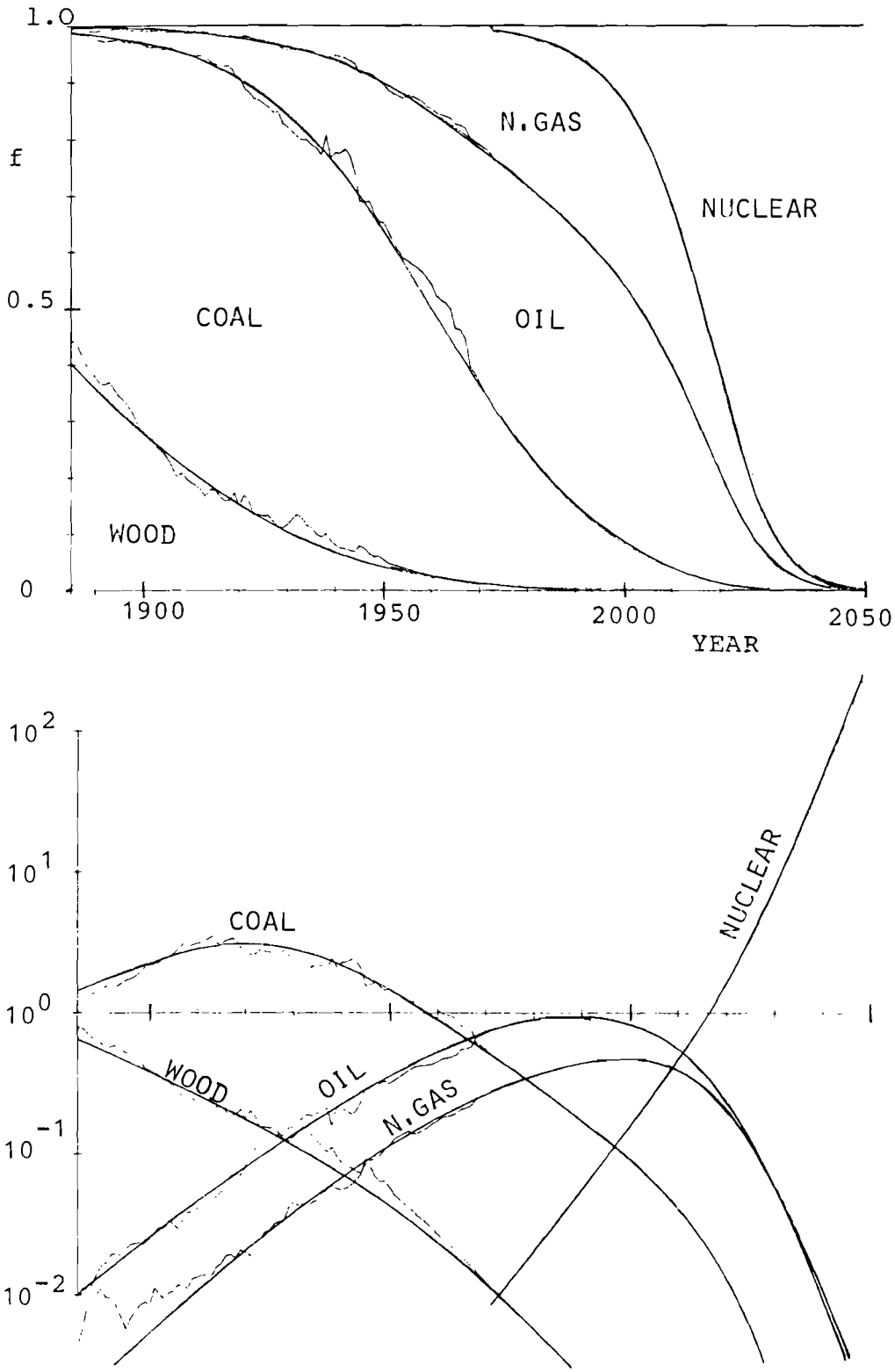


Figure 21. Forecast of market penetration by nuclear energy--economic assessment from [20].

in this example, indicates that this question may be not very critical.

The meaning of the curves presented in Figures 19a,b and 20a,b corresponds to the probabilistic interpretation of the output of the deterministic model given in this section, under the assumption that the parameter values c_{54}, a_{54} (in the first case) and c_{53}, a_{53} (in the second case) are certain. This is, of course, not true. If we wanted to be more objective it would be possible to construct a "subjective" probability distribution for these uncertain parameters, based on all available assessments and opinions, and to project this uncertainty into the future as outlined in this section. In this way the probability distribution of the future process output conditioned on the present knowledge could be obtained.

8. CONCLUDING REMARKS

Any scientific approach to the problem of forecasting cannot be anything else than drawing conclusions on the basis of certain assumptions. It is possible to check whether these assumptions have been fulfilled in the past but, strictly speaking, it never can be guaranteed that they will not be violated in the future. Employing mathematics in this study, it was possible to base the forecast on a few simple and well understandable assumptions and to maintain consistency of reasoning even in rather complex situations. This makes it possible to reduce the discussion about the possible future to a criticism of the basic assumptions.

Perhaps the most important assumption, on which the presented theory is based, is that in the long term a particular technology has to live and grow on its own account, i.e. that the mean value of the external capital flow is zero. The assumption that the mean values of certain economic characteristics do not change in time restricts the validity of the model to the situations when the technology starting to penetrate the market is already well established. The model also cannot predict the birth of a new technology. It must be introduced into the model exogenously.

The penetration of a market by new technologies is a very complex interplay between producers and consumers. This study emphasizes the macroeconomic view on the producer side. The consumer side, as discussed at the beginning of Section 4, is reflected in the correction by which the total production costs have to be reduced in order to respect the difference in market price the average consumer is ready to pay for the higher quality of satisfaction of his need. In some cases this correcting term cannot be considered as stationary and can be influenced by advertisement and/or by official propaganda. In these cases the consumer side and the spread of information should be considered in more detail.

In societies with planned economies the mechanism of an open market is replaced by economic balances and decisions made by planning institutions and committees. However, not even a planned economy can afford to support a loser without special reasons. The planners also have to respect the social demand in order to ensure a fluent and regular distribution of products but they can control the substitution process, by setting taxes and different prices (both can be reflected in the model in total production costs), in order to achieve some goals. It is believed that the model developed in this study could serve as a planning tool for these purposes.

Speaking about possible control of the substitution process another important point has to be mentioned. As pointed out in Section 4 no technology can start from zero without external financial help. The magnitude of the initial external investment actually determines the initial conditions for the model and may considerably accelerate (or delay if it is too small) the substitution process, especially when the new technology is profitable but requires high investments. This is clearly seen from the sensitivity analysis performed in Example 2.

In Example 5 the forecast of penetration of the energy market by nuclear energy, based on its economic assessment, has been given. Recently, much attention has been paid to the

question whether such a technological development would not be too risky. Many other technologies are risky, too. (Consider, for instance, the automobile -- one of the main killers of mankind.) Whether a risk has been taken or not is also an economical question. The presented model could help to evaluate the economic loss the society should accept in order to avoid the risk. In this way it would be possible to base the decision on a more objective (and less emotional) basis.

ACKNOWLEDGMENTS

The author wishes to express his thanks to several members of the international scientific staff at IIASA.

The main impulse for this study has been given by the work of C. Marchetti with whom the author has been in continuing contact. The world energy data have been collected by N. Nakicenovic who also performed the first calculations. Most of the programming effort and numerical experiments have been done with the assistance of M. Valasek of the IIASA computer services group.

Last, but not least, the author's gratitude goes to Prof. Haefele, the leader of IIASA Energy Project, for his support and encouragement.

Appendix A. Mathematical theorems and proofs

To simplify the proofs of theorems used in the main text four known Lemmas will be stated first.

Lemma 1. Let A and D be nonsingular square matrices, maybe of different orders, and B and C matrices of appropriate dimensions. Then the following relations for determinants hold:

$$\begin{vmatrix} A & C \\ B & D \end{vmatrix} = |A||D - BA^{-1}C| = |D||A - CD^{-1}B| \tag{A.1}$$

$$|A - CD^{-1}B| = \frac{|A|}{|D|}|D - BA^{-1}C| \tag{A.2}$$

Proof: See, e.g., Rao [18], supplement to Chapter 1b.

Lemma 2. Let A be a nonsingular square matrix, b a vector and γ a scalar. Then

$$\left[A + b\frac{1}{\gamma}b^T\right]^{-1} = A^{-1} - \frac{1}{\gamma + b^TA^{-1}b}A^{-1}bb^TA^{-1} \tag{A.3}$$

$$\left|A + b\frac{1}{\gamma}b^T\right| = |A|\left(1 + \frac{1}{\gamma}b^TA^{-1}b\right) . \tag{A.4}$$

Proof. To prove (A.3) multiply both its sides by $(A + b\frac{1}{\gamma}b^T)$. The second equality (A.4) is a special case of (A.2).

Lemma 3. Let M be a positive definite matrix of dimension $(\nu \times \nu)$. Then

$$\int_{R^\nu} \exp\{-\tfrac{1}{2}(x - c)^T M(x - c)\}dx = (2\pi)^{\frac{\nu}{2}}|M|^{-\frac{1}{2}} \tag{A.5}$$

Proof. See, e.g., Anderson [17] §2.3.

Lemma 4. Let S_Ω be a space of all positive definite matrices Ω of dimension ν and let ϕ be also positive definite. Then

$$\int_{S_\Omega} |\Omega|^{\frac{\mu}{2}} \exp\{-\tfrac{1}{2}\operatorname{tr}(\Omega\phi)\}d\Omega$$

$$= (2^{\mu+\nu+1}\pi)^{\frac{\nu-1}{2}} \prod_{j=1}^{\nu} \Gamma\left(\frac{\mu+\nu+2-j}{2}\right) \cdot |\phi|^{-\frac{\mu+\nu+1}{2}} , \qquad (A.6)$$

where $\Gamma(\cdot)$ is the gamma-function.

Proof. See, e.g., DeGroot [14] §5.5.

Theorem 1. The transformation between two ν-dimensional real vectors $\{e_i : i = 1,2,\ldots,\nu\}$ defined by relations

$$\ell n\, f_i - a_i\, \ell n\, f_{\nu+1} + k_i = e_i , \qquad i = 1,2,\ldots,\nu \qquad (A.7)$$

$$f_{\nu+1} = 1 - \sum_{i=1}^{\nu} f_i \qquad (A.8)$$

$$0 < f_i < 1 , \quad \forall i ,$$

where all a_i are positive and real, is regular (one-to-one) and its Jacobian is

$$J = \frac{\sum_{i=1}^{\nu} f_i a_i + f_{\nu+1}}{\prod_{i=1}^{\nu+1} f_i} . \qquad (A.9)$$

Proof. Notice that the transformation is continuous. To prove its regularity it is sufficient to show that all partial derivatives $\partial e_i/\partial f_j$ are positive. From (A.7) and (A.8) we have

$$j \neq i \quad : \quad \frac{\partial e_i}{\partial f_j} = \frac{a_i}{f_{\nu+1}} > 0 \tag{A.10}$$

$$\frac{\partial e_i}{\partial f_i} = \frac{1}{f_i} + \frac{a_i}{f_{\nu+1}} > 0 \tag{A.11}$$

and the Jacobian can be expressed as follows:

$$J = \left| \begin{bmatrix} f_1^{-1} & & & 0 \\ & f_2^{-1} & & \\ & & \ddots & \\ 0 & & & f_\nu^{-1} \end{bmatrix} + \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix} \frac{1}{f_{\nu+1}} \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a \end{bmatrix}^T \right| .$$

Application of (A.2) with $A^{-1} = \text{diag}\,(f_1, f_2, \ldots, f_\nu)$, $C^T = [1,1,\ldots,1]$, $D = f_{\nu+1}$ and $B = [a_1, a_2, \ldots, a_\nu]$ proves (A.9).

Theorem 2. Let $f(\Omega)$ *be a scalar function of a positive semidefinite matrix* Ω *of dimension* ν *defined by*

$$f(\Omega) = |\Omega|^{\frac{\rho}{2}} e^{-\frac{1}{2}\text{tr}(\Omega D)} , \tag{A.12}$$

where D *is positive definite. The maximum of* $f(\Omega)$ *is taken on at*

$$\Omega = \rho D^{-1} \tag{A.13}$$

and the maximum is

$$f(\rho D^{-1}) = \left(\frac{\rho}{e}\right)^{\frac{\rho\nu}{2}} |D|^{-\frac{\rho}{2}} . \tag{A.14}$$

Proof. See Anderson [17] §3.2.

Theorem 3. Let H *be a symmetric nonsingular* $(n \times n)$ *matrix and*

$$A = [I , -a] , \tag{A.15}$$

where I *is an identity matrix of dimension* n-1 *and* a *is a column vector. Then the following equality holds*

$$|A\,H\,A^T| \;=\; |H|\;(\tilde{a}^T H^{-1}\,\tilde{a}) \tag{A.16}$$

where $\tilde{a}$ *is the vector* a *extended by* 1.

$$\tilde{a} \;=\; \begin{bmatrix} a \\ 1 \end{bmatrix} \quad .$$

Proof. Consider the matrix H and its inverse $G = H^{-1}$ partitioned in the following way

$$H \;=\; \begin{bmatrix} H_r & h \\ h^T & \eta \end{bmatrix} \quad , \qquad G \;=\; \begin{bmatrix} G_r & g \\ g^T & \gamma \end{bmatrix} \quad , \tag{A.17}$$

where h,g are vectors and η,γ scalars. From the equality HG = I the following relations are obtained.

$$G_r \;=\; \left(H_r \;-\; h\,\frac{1}{\eta}\,h^T\right)^{-1} \tag{A.18}$$

$$\gamma \;=\; \frac{1}{\eta} + \frac{1}{\eta^2}\,h^T G_r\,h \tag{A.19}$$

$$g \;=\; -G_r\,h\,\frac{1}{\eta} \quad . \tag{A.20}$$

Using the definition (A.15) of the matrix A we have

$$|A\,H\,A^T| \;=\; \left|\,[I\,,\,-a]\begin{bmatrix} H_r & h \\ h^T & \eta \end{bmatrix}\begin{bmatrix} I \\ -a^T \end{bmatrix}\right|$$

$$\;=\; |H_r \;-\; h\,a^T \;-\; a\,h^T \;+\; a\,\frac{1}{\eta}\,a^T|$$

$$\;=\; |\,(H_r \;-\; h\,\frac{1}{\eta}\,h^T) \;+\; (a \;-\; \frac{1}{\eta}\,h)\,\eta\,(a - \frac{1}{\eta}\,h)^T\,| \quad . \tag{A.21}$$

Making use of (A.4) we obtain

$$|A\,H\,A^T| = \left|H_r - h\frac{1}{\eta}h^T\right|\left(1 + \eta\left(a - \frac{1}{\eta}h\right)^T\left(H_r - h\frac{1}{\eta}h^T\right)^{-1}\left(a - \frac{1}{\eta}h\right)\right) \quad (A.22)$$

and reversed application of (A.1) gives

$$\left|H_r - h\frac{1}{\eta}h^T\right| = \frac{1}{\eta}|H| \ .$$

Considering also (A.18), (A.19) and (A.20) we can continue in rearranging of (A.22) as follows

$$\begin{aligned}\left|A\,H\,A^T\right| &= |H|\left(\frac{1}{\eta} + \left(a - \frac{1}{\eta}h\right)^T G_r\left(a - \frac{1}{\eta}h\right)\right)\\ &= |H|\left(\frac{1}{\eta} + \frac{1}{\eta^2}\,h^T G_r h - \frac{1}{\eta}h^T G_r a - a^T G_r\,h\frac{1}{\eta} + a^T G_r a\right)\\ &= |H|\left(\gamma + g^T a + a^T g + a^T G_r a\right) \ . \end{aligned} \quad (A.23)$$

However, the last expression in (A.23) can be written also in this way

$$|A\,H\,A^T| = |H|\ [a^T, 1]\begin{bmatrix} G_r & g \\ g^T & \gamma \end{bmatrix}\begin{bmatrix} a \\ 1 \end{bmatrix}$$

which completes the proof.

Theorem 4. Let $H_{(N+1)}$ *be a* $(n \times n)$*-matrix defined similarly to (6.88)*

$$H_{(N+1)} = \phi_{(N+1)} - b_{(N+1)}\,\tau_{(N+1)}\,b^T_{(N+1)} \ , \quad (A.24)$$

where, according to (6.77) and (6.80)

$$\phi_{(N+1)} = \sum_{k=2}^{N+1}\frac{1}{T_{(k)}}\,x_{(k)}\,x^T_{(k)} \quad (A.25)$$

$$T_{(k)} = t_k - t_{k-1} \tag{A.26}$$

$$\tau_{(N+1)} = t_{N+1} - t_1 \quad , \tag{A.27}$$

and $b_{(N+1)}$ *is a* n*-vector the components of which are defined according to (6.87)*

$$b_{(N+1)i} = \frac{1}{\tau_{(N+1)}} \ell n \frac{f_{(N+1)i}}{f_{(1)i}} \quad , \quad \forall i \quad . \tag{A.28}$$

Then the following recursive relations hold:

$$\tau_{(N+1)} = \tau_{(N)} + T_{(N+1)} \tag{A.29}$$

$$b_{(N+1)} = \frac{\tau_{(N)}}{\tau_{(N+1)}} \left(b_{(N)} + \frac{x_{(N+1)}}{\tau_{(N)}} \right) \tag{A.30}$$

$$H_{(N+1)} = H_{(N)} + \frac{1}{\theta_{(N+1)}} (x_{(N+1)} - T_{(N+1)} b_{(N)}) (x_{(N+1)} - T_{(N+1)} b_{(N)})^T \tag{A.31}$$

$$\theta_{(N+1)} = \frac{T_{(N+1)} \tau_{(N+1)}}{\tau_{(N)}} \tag{A.32}$$

$$H^{-1}_{(N+1)} = H^{-1}_{(N)} - \frac{1}{\gamma_{(N+1)}} d_{(N+1)} d^T_{(N+1)} \tag{A.33}$$

$$d_{(N+1)} = (x_{(N+1)} - T_{(N+1)} b_{(N)}) \tag{A.34}$$

$$\gamma_{(N+1)} = \theta_{(N+1)} + d^T_{(N+1)} H^{-1}_{(N)} d_{(N+1)} \tag{A.35}$$

$$|H_{(N+1)}| = \theta_{(N+1)} |H_{(N)}| \gamma_{(N+1)} \quad . \tag{A.36}$$

Proof. Relation (A.29) directly follows from the definitions of $\tau_{(N+1)}$ (A.27) and $T_{(k)}$ (A.26).

The i-th component of the vector $b_{(N+1)}$ (A.28) can be written as follows:

$$b_{(N+1)i} = \frac{1}{\tau_{(N+1)}} \left(\ell n \frac{f_{(N)i}}{f_{(1)i}} + \ell n \frac{f_{(N+1)i}}{f_{(N)i}} \right)$$

$$= \frac{\tau_{(N)}}{\tau_{(N+1)}} \left(b_{(N)i} + \frac{1}{\tau_{(N)}} x_{(N+1)i} \right) ,$$

which proves (A.30).

Using (A.25) and (A.30) the matrix $H_{(N+1)}$ (A.24) can be expressed in the following way:

$$H_{(N+1)} = \phi_{(N)} + \frac{1}{T_{(N+1)}} x_{(N+1)} x^T_{(N+1)}$$

$$- \left(b_{(N)} + \frac{x_{(N+1)}}{\tau_{(N)}} \right) \frac{\tau^2_{(N)}}{\tau_{(N+1)}} \left(b_{(N)} + \frac{x_{(N+1)}}{\tau_{(N)}} \right)^T$$

and after a simple rearrangement, (A.31) is obtained.

Application of (A.3) and (A.4) to (A.31) gives (A.33) and (A.36), respectively, which completes the proof.

Theorem 5. Let $L\left(c,a,\Omega,y_{(1)}^{(M)}\right)$ *be the likelihood function defined by (6.81) for* $M = N$, *where* $c \in S_c$ *is a* $(n-1)$*-vector and* $\Omega \in S_\Omega$ *a* $(n-1) \times (n-1)$*-matrix. The integral of this function taken over the* $(n-1)$*-dimensional space* $S_c \equiv R^{n-1}$ *and over the space* S_Ω *of all positive definite matrices of dimension* $(n-1) \times (n-1)$ *is*

$$\kappa\left(y_{(1)}^{(M)},a\right) = \int_{S_c} \int_{S_\Omega} L\left(c,a,\Omega,y_{(1)}^{(M)}\right) dc\, d\Omega =$$

cont./...

$$= \left(\frac{2^{n-1}}{\pi^{M-\frac{n}{2}} \prod_{k=2}^{N} T_{(k)}} \right)^{\frac{n-1}{2}} \prod_{j=1}^{n-1} \Gamma\left(\frac{M+n-j}{2}\right) \cdot \frac{\prod_{k=2}^{M} (f_{(k)}^T \tilde{a})}{\prod_{k=2}^{M} \Pi(f_{(k)})}$$

$$\times \left| A H_{(M)} A^T \right|^{-\frac{M+n-1}{2}} \qquad \text{(A.37)}$$

where $\Gamma(\cdot)$ *means gamma-function and* $|\cdot|$ *means determinant. For* $M = N + 1$ *the determinant in (A.37) can be expressed as follows:*

$$\left| A H_{(N+1)} A^T \right| = \left| A H_{(N)} A^T \right| \cdot \left(1 + \hat{e}_{(N+1)}^T \psi_{(N)}^{-1} \hat{e}_{(N+1)} \right) \qquad \text{(A.38)}$$

$$\hat{e}_{(N+1)} = A x_{(N+1)} - \hat{c}_{(N)}^T \phi_{(N+1)} \qquad \text{(A.39)}$$

$$\psi_{(N)} = T_{(N+1)} \frac{\tau_{(N+1)}}{\tau_{(N)}} A H_{(N)} A^T ,$$

where $x_{(N+1)}$ *is the vector introduced by (6.61) and (6.62), and* $\hat{c}_{(N)}$ *is the maximum likelihood estimates (6.90) of the parameter vector* c*, available at time* t_N*. Hence*

$$\frac{\kappa\left(y_{(1)}^{(N+1)}, a\right)}{\kappa\left(y_{(1)}^{(N)}, a\right)} = \frac{\Gamma\left(\frac{N+n-1}{2}\right)}{\Gamma\left(\frac{N}{2}\right)} \frac{\pi^{\frac{n-1}{2}}}{|\psi_{(N)}|} \frac{f_{(N+1)}^T \tilde{a}}{\Pi(f_{(N+1)})}$$

$$\times \frac{1}{\left(1 + \hat{e}_{(N+1)}^T \psi_{(N)}^{-1} \hat{e}_{(N+1)} \right)^{\frac{N+n}{2}}} . \qquad \text{(A.40)}$$

Proof. To proof (A.37) rearrange the exponent in the likelihood function (6.81) according to (6.85) and apply Lemma 3 first and Lemma 4 afterwards.

Using (A.31) the determinant on the left-hand side of (A.38) can be expressed as follows:

$$\left|A\,H_{(N+1)}\,A^T\right|$$

$$= \left|A\,H_{(N)}\,A^T + A(x_{(N+1)} - T_{(N+1)}\,b_{(N)})\frac{1}{\theta_{(N+1)}}(x_{(N+1)} - T_{(N+1)}\,b_{(N)})^T A^T\right| .$$

According to (6.90)

$$A\,b_{(N)} = \hat{c}_{(N)}$$

and consequently

$$\left|A\,H_{(N+1)}\,A^T\right| = \left|A\,H_{(N)}\,A^T + \hat{e}_{(N+1)}\frac{1}{\theta_{(N+1)}}\hat{e}^T_{(N+1)}\right| ,$$

where $\hat{e}_{(N+1)}$ has been introduced by (A.39). Now apply (A.4) to prove equality (A.38).

Appendix B. An optimization method

In Section 6 we had to find the vector a maximizing the function

$$\lambda(a) = \frac{\prod_{k=2}^{N} f_{(k)}^T \tilde{a}}{\left[\tilde{a}^T H^{-1} \tilde{a}\right]^{\frac{N-1}{2}}} , \tag{B.1}$$

where

$$\tilde{a} = \begin{bmatrix} a \\ 1 \end{bmatrix} . \tag{B.2}$$

In the sequel we shall propose a simple and effective numerical method which serves this purpose.

As already pointed out, the value of the function λ does not change when a in (B.1) is substituted for

$$\bar{a} = \frac{1}{\gamma} \tilde{a} \tag{B.3}$$

where γ is any nonzero number, not necessarily constant. Let us choose γ as follows:

$$\gamma = k(\tilde{a}^T H^{-1} \tilde{a})^{\frac{1}{2}}$$

where k is an arbitrary constant. Then the function (B.1) gets a simple form

$$\lambda(\bar{a}) = \prod_{k=2}^{N} f_{(k)}^T \bar{a} \tag{B.4}$$

but the vector $\bar{a}$, the last component of which is no longer fixed to be one, is restricted by the condition

$$\bar{a}^T H^{-1} \bar{a} = \frac{1}{k^2} . \tag{B.5}$$

In this way the optimization problem is transformed into the problem of finding a point $\bar{a}^*$ on the ellipsoid (B.5) at

which the function (B.4) is maximal. In other words, we have to find the point $\bar{a}^*$ at which the direction of the gradient with respect to the ellipsoid

$$g_e(\bar{a}) = \text{grad}\ \{\bar{a}^T H^{-1} \bar{a}\} = H^{-1}\bar{a} \tag{B.6}$$

and the direction of the gradient with respect to the function $\lambda(\bar{a})$

$$g_\lambda(\bar{a}) = \text{grad}\ \{\lambda(\bar{a})\}$$

$$g_\lambda(\bar{a})_i = \prod_{m=2}^{N} f_{(m)}^T \bar{a} \cdot \sum_{k=2}^{N} \frac{f_i}{f_{(k)}^T \bar{a}} \tag{B.7}$$

coincide, as shown in Figure B1. Notice that only directions are significant.

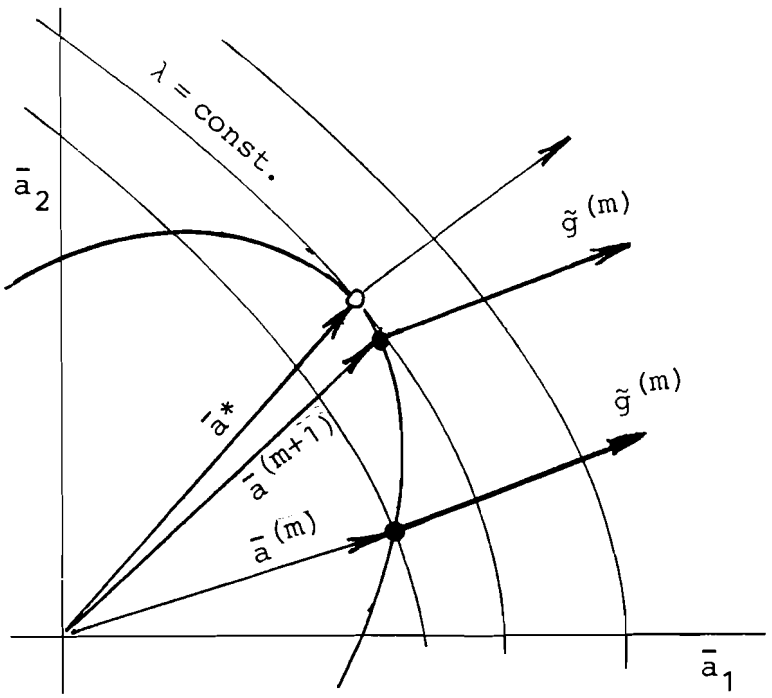


Figure B1. Optimization algorithm.

The optimum point $\bar{a}^*$ can be found using the following

A l g o r i t h m AMLE:

1. Choose any starting point $\bar{a}^{(m)}$, m = 0, say $\bar{a}^{(0)} = [1,1,\ldots,1]^T$.

2. Calculate the direction of the gradient with respect to the function λ

$$g_\lambda(a^{(m)}) \propto \tilde{g}^{(m)}$$

$$\tilde{g}_i^{(m)} = \sum_{k=2}^{N} \frac{f_{(k)i}}{f_{(k)}^T \bar{a}^{(m)}} \quad . \tag{B.8}$$

3. Find the next point $a^{(m+1)}$ in which the gradient $g_e(a^{(m+1)})$ with respect to the ellipsoid (B.6) has the same direction as the gradient $g_\lambda(a^{(m)})$.

$$H^{-1}\bar{a}^{(m+1)} = \tilde{g}^{(m)}$$

$$\bar{a}^{(m+1)} = H\tilde{g}^{(m)} \quad . \tag{B.9}$$

4. Normalize the vector $\bar{a}^{(m+1)}$, say to unit length, and repeat 2 until the vectors $\bar{a}^{(m)}$ and $\bar{a}^{(m+1)}$ coincide with given precision.

The FORTRAN - subroutine realizing this algorithm can be found in Appendix C.

No convergence proof is available for the optimization algorithm described but it never failed on a number of examples and appeared to be very fast.

Appendix C. FORTRAN Subroutines

```
      subroutine penetr(n,c,al,ro,f0,t,f)
c
c   function          - market penetration - prediction
c                       of market shares
c   usage             - call penetr(n,c,al,ro,f0,t,f)
c   parameters      n - number of competitors , max.10
c                   c - n-vector of production costs;
c                       as the result depends only on differencies
c                       between costs any number can be substracted
c                       from all c's.
c                  al - n-vector of specific investments;
c                       only nonzero components are allowed
c                  ro - growth rate factor of total production of all
c                       competitors
c                  f0 - n-vector of initial market shares at t=0;
c                       the sum of all f0's must be equal to 1
c                   t - the value of time for which the prediction
c                       is calculated; t can be negative.
c                   f - output : n-vector of predicted market shares;
c                       the sum of all elements of f is equal to 1.00
c   precision         - 3 decimal digits behind the decimal point
c                       of all f's; if higher precision is required
c                       see c1
c
      dimension c(1),al(1),f0(1),f(1)
      dimension a(10),beta(10)
c                       calculate alm
      alm=0.
      sum=0.
      do 5 i=1,n
      ali=al(i)
      alm=alm+1./ali
    5 sum=sum+1./(ali*ali)
      alm=alm/sum
c                       calculate a,beta and initial psi
      sum=0.
      do 10 i=1,n
      a(i)=alm/al(i)
      beta(i)=(c(i)+al(i)*ro)/alm*t
   10 sum=sum+f0(i)*exp(-beta(i))
      psi=-alog(sum)
c                       iterative solution of eqn. xi(psi)=0
   15 xi=-1.
      dxi=0.
      do 20 i=1,n
      f(i)=f0(i)*exp((psi-beta(i))*a(i))
      xi=xi+f(i)
   20 dxi=dxi+a(i)*f(i)
      if(abs(xi).lt.0.5e-3) go to 25
c1                      if higher precision is required change
c                       0.5e-3 to 0.5e-d where d is the number
c                       of decimal digits behind the decimal point
c                       of f's which have to be guaranteed
      psi=psi-xi/dxi
      go to 15
   25 return
      end
%
```

```
      subroutine mlest(n,ir,t,f,kf,kl,cir,air,r,clir,rl,
     *h,b,tau,it,itest)
c
c  function     - market penetration - maximum likelihood estimation
c                 of model parameters
c
c  usage        - call mlest(n,ir,t,f,kf,kl,cir,air,r,clir,rl,
c                *h,b,tau,it,itest)
c                 requires the subroutine amle
c  parameters   -
c  input:     n - number of competitors,max.7
c             ir - index of the reference competitor
c              t - vector of time points for which the market
c                 shares are given
c              f - matrix of market shares, f(k,i)=f(t(k),i),
c                 sum f(k,i):i=1,n  must be 1.
c             kf - first row in t and f which is considered in estimation
c             kl - last row in t and f considered,kf.lt.kl.le.100
c  output:  cir - ml-estimate of c(i,ir):i=1,n; c(ir,ir)=0.
c           air - ml-estimate of a(i,ir):i=1,n; a(ir,ir)=1.
c             r - ml-estimate of covariance matrix r
c          clir - ml-estimate of c(i,ir) under assumption a(i,ir)=1
c                 for all i.
c            rl - ml-estimate of covariance matrix r under assumption
c                 a(i,ir)=1 for all i.
c             h - auxiliary matrix required for bayesian forecasting
c             b - vector for bayesian forecasting
c           tau - scalar parameter for bayesian forecasting
c            it - number of iterations performed in calculation
c                 of air. maximum it is set to 50. if it=50
c                 the iteration has not been completed.
c         itest - regular case itest=1. if itest=0 then the calcu-
c                 lation could not be completed because one or more
c                 a's are zero, in that case cir and r are set to zero
c                 and a,i=1,n is given out on the place of air.
c
      dimension t(100),f(100,7),h(7,7),b(7),a(7),
     &air(7),cir(7),clir(7),r(7,7),rl(7,7)
c
      m=kl-kf
      tau=t(kl)-t(kf)
      ks=kf+1
c                              calculate vector b
      do 12 i=1,n
   12 b(i)=alog(f(kl,i)/f(kf,i))/tau
c                              calculate matrix h
      do 16 i=1,n
      do 16 j=1,i
      sum=0.
      do 14 k=ks,kl
      l=k-1
      dt=t(k)-t(l)
      sum=sum+(alog(f(k,i)/f(l,i))/dt-b(i))
     &*(alog(f(k,j)/f(l,j))/dt-b(j))*dt
   14 continue
      h(i,j)=sum
```

```
   16 h(j,i)=sum
c                              calculate estimates of a
      call amle(n,f,ks,kl,h,a,it,itest)
c
      if(itest.gt.0)go to 20
      do 18 i=1,n
      cir(i)=0.
      air(i)=a(i)
      do 18 j=1,n
   18 r(i,j)=0.
      go to 28
c                              calculate vectors air and cir
   20 do 22 i=1,n
      air(i)=a(ir)/a(i)
   22 cir(i)=b(ir)-air(i)*b(i)
c                              calculate covariance matrix r
      do 26 i=1,n
      do 26 j=1,i
      r(i,j)=(h(i,j)-h(i,ir)/air(j)-h(ir,j)/air(i)+
     &h(ir,ir)/air(i)/air(j))/m
   26 r(j,i)=r(i,j)
c                              calculate vector clir
   28 do 30 i=1,n
   30 clir(i)=b(ir)-b(i)
c                              calculate covariance matrix rl
      do 32 i=1,n
      do 32 j=1,i
      rl(i,j)=(h(i,j)-h(i,ir)-h(ir,j)+h(ir,ir))/m
   32 rl(j,i)=rl(i,j)
      return
      end
```

```
      subroutine amle(n,f,ks,kl,h,a,it,itest)
c
c  function     - market penetration -maximization of likeli-
c                 hood function with respect to vector a
c  usage        - call amle(n,f,ks,kl,h,a,it,itest)
c  parameters   - see subroutine mlest
c  remark       - output vector a is normalized to unit length.
c
      dimension a(7),ap(7),g(7),f(100,7),h(7,7)
      it=0
      xn=n
      sum=1./sqrt(xn)
      do 101 i=1,n
  101 ap(i)=sum
c
  102 itest=1
      do 103 i=1,n
  103 g(i)=0.
      do 105 k=ks,kl
      sum=0.
      do 104 i=1,n
  104 sum=sum+f(k,i)*ap(i)
      do 105 i=1,n
  105 g(i)=g(i)+f(k,i)/sum
      ra=0.
      do 108 i=1,n
      sum=0.
      do 106 k=1,n
  106 sum=sum+h(i,k)*g(k)
      if(sum.gt.0.)go to 107
      itest=0
      sum=0.
  107 a(i)=sum
  108 ra=ra+sum*sum
      ra=sqrt(ra)
      sum=0.
      do 109 i=1,n
      a(i)=a(i)/ra
      sum=sum+(a(i)-ap(i))**2
  109 ap(i)=a(i)
c
      it=it+1
      if(sum.lt.1.e-6)go to 110
      if(it.lt.50)go to 102
  110 return
      end
```

REFERENCES

[1] Marchetti, C., *Primary Substitution Models*, internal paper, June 1975, see also Volume I of this report.

[2] Marchetti, C., *On Strategies and Fate*, in W. Häfele et al. *Second Status Report on the IIASA Project on Energy Systems 1975*, RR-76-1, International Institute for Applied Systems Analysis, Laxenburg, Austria, 1976, see also Volume I of this report.

[3] Fisher, J.C., and R.H. Pry, *A Simple Substitution Model of Technological Change*, Report 70-C-215, General Electric Company, Research and Development Center, Schenectady N.Y., Technical information series, June, 1970; see also *Technological Forecasting and Social Change*, 3, (1971), 75-88.

[4] Pry, R.H., *Forecasting the Diffusion of Technology*, Report 73CRD220, General Electric Company, Corporate Research and Development, Schenectady N.Y., Technical information series, July 1973.

[5] Blackman, A.W., A Mathematical Model for Trend Forecasts, *Technological Forecasting and Social Change*, 3, (1972), 441-452.

[6] Blackman, A.W., The Market Dynamics of Technological Substitution, *Technological Forecasting and Social Change*, 6, (1974), 41-63.

[7] Floyd, A., *Trend Forecasting: A Methodology for Figure of Merit*, in Proceedings of the First Annual Technology and Management Conference, J. Bright, ed., Prentice-Hall, Englewood Cliffs, New Jersey, 1968.

[8] Sharif, M.N., and Ch. Kabir, A Generalized Model for Forecasting Technological Substitution, *Technological Forecasting and Social Change*, 8, (1976), 353-364.

[9] Mansfield, E., Technical Change and the Rate of Imitation, *Econometrica*, 29, 4 (Oct. 1961).

[10] Bundgaard-Nielsen, M., The International Diffusion of New Technology, *Technological Forecasting and Social Change*, 8, (1976), 365-370.

[11] Andress, F.J., The Learning Curve As a Production Tool, *Harvard Business Review*, 32 (1954), pp. 87-97.

[12] Nakicenovic, N., private communication, see also Volume I of this report.

[13] De Finetti, B., *Theory of Probability: a critical introductory treatment*, Volume 1, Wiley, New York, 1974.

[14] DeGroot, M.H., *Optimal Statistical Decisions*, McGraw-Hill, New York, 1970.

[15] Raiffa, H., Schleifer, R., *Applied Statistical Decision Theory*, Harvard University, Boston, 1961.

[16] Lindley, D.V., *The Future of Statistics - A Bayesian 21'st Century*, in Proceedings of the Conference on Directions for Mathematical Statistics, Canada, 1974, *Supp.Adv.Appl.Prob.* 7 (1975), 106-115.

[17] Anderson, T.V., *An Introduction to Multivariate Statistical Analysis*, John Wiley, 1958.

[18] Rao, C.R., *Linear Statistical Inference and Its Applications*, John Wiley, 1965.

[19] *Einsatzmöglichkeiten neuer Energiesysteme, Programmstudie "Sekundärenergiesysteme"*, Eine Information des Bundesministers für Forschung und Technologie, Teil I: Bedarfsanalyse und Strom, Bonn 1975.

[20] Michaelis, H., *Status der Wettbewerbsfähigkeit der Kernenergie*, Reaktortagung '76, Düsseldorf, 1976.

The Regularity of Market Penetration

F. Fleck

1. INTRODUCTION AND BRIEF REVIEW OF THE LITERATURE

Although innovations may vary greatly, their market penetration processes show a uniform picture. It is therefore natural to assume that these processes have basic characteristics in common. Most studies published on this subject try to describe these processes in mathematical models and to plot the available data as accurately as possible. The most accurate results were obtained with the logistic growth function.

In Volume I of this report it was shown that the primary energy substitution processes can also be described very accurately by logistic functions. That was the impulse for this paper which is part of a Ph.D. thesis written at the University of Karlsruhe. The aim of this study is to analyze the penetration pocess, to determine the facts giving rise to the regularities of these processes, and to organize these facts in a suitable formal model. Subsequently, the quality of the model is to be checked by applying it to some practical examples.

The first attempts at a mathematical definition of such penetration or expansion processes were made by biologists. Above all Pearl (1925) used the logistic function, sometimes called Pearl function, to precisely describe the population growth in the US and the reproduction of yeast cells. The logistic function used by Pearl has the following form:

$$y = \frac{k}{1+e^{-ax-b}} \quad , \tag{1}$$

where y indicates penetration as a function of x (x can be often

interpreted as time t). a,b,k are parameters which have to be determined, with k representing the absolute limit of saturation. Other biologists found that epidemics spread in a similar way. They used stochastic formulations rather than deterministic ones in order to account for the indefiniteness of such developments. Barlett (1946), Bailey (1950), and Haskey (1954) should be mentioned in this context.

Geographers like Hägerstrand (1952, 1967) and Hudson (1969), who studied mainly the spatial aspects of the spreading of innovations, greatly contributed to the explanation of penetration processes. Hägerstrand was one of the first to realize that the diffusion of the information on the innovation plays an important role when the product penetrates the market.

An impressive starting point for many research studies has been a study by Griliches (1957) which deals with the penetration of a new seed (hybrid corn) into the market of US farmers. In this case we can speak of a substitution process, since the innovation (new seed) did not enter an empty market but had to displace the established product (ordinary seeds). Using the logistic function for this systematic and comprehensive study, Griliches plotted the substitution process very accurately.

Mansfield (1961) contributed a fundamental study of the industrial sector. Using the logistic function, he studied the market penetration of twelve innovations in four industrial branches, concentrating on the penetration speed of innovations. This work gave rise to a wide range of similar studies. One of them is a revision of Mansfield's model by Blackman (1974). He no longer described market penetration in absolute numbers but expressed it in relative quantities of market shares.

The model by Bass (1969) for the first time included aspects of behavioral theory and a differentiation between consumers. How suitable the logistic function is for describing such penetration processes is clearly demonstrated by the studies of Fisher and Pry (1970), and Pry (1973). They used Pearl's

function but introduced a relative saturation limit (market share).

On the basis of these studies Marchetti found that the primary energy market can also be described in this way. He had to modify the basis in such a way as to assure that more than two products can be studied. In the first part of this volume Peterka extended this approach and theoretically substantiated this empirical analysis. His macroeconomic analysis in the perspective of production and capital led to a generalization of the Fisher-Pry model.

In this context it should also be mentioned that the logistic function is not the only way of exactly describing such penetration processes. Stapleton (1976) proved with the Fisher-Pry model that accumulated normal distribution also supplies exact values. This result is confirmed by an early study by Winsor (1932), which shows that among different symmetrical growth curves the logistic function and the accumulated normal distribution clearly give the most accurate results.

The widespread use of the logistic curve is probably due to the fact that the basic differential equation can be easily interpreted and that the assessment of parameters is relatively simple. An important characteristic both curves have in common is the S-like shape of all penetration processes. The reasons for this property of the functions are discussed in the following section.

2. ANALYSIS OF THE PENETRATION PROCESS

In most models the S-shape of the curve is taken for granted and the decisive factors are not analyzed. Sahal's (1976) study is an exception. From his attempts of systematizing the reasons it becomes clear that not only economic factors but also sociological and psychological factors play an important role. In general, however, the reasons given do not follow from a general analysis but result from the special intentions of each model builder.

Thus numerous and relevant reasons are offered for an appropriate interpretation of the model parameters, but there is hardly any reference to the facts all of these processes have in common. For detailed analysis, however, one has to abandon the macroscopic approach towards such processes and change over to an individual microscopic analysis, since it is the individual in an economic and social environment who decides on the adoption or rejection of an innovation, i.e. the market penetration of an innovation is controlled by the decisions of individuals. The individual's decisions are mainly influenced by the diffusion of the information available on the innovation, and they are actually a result of the decision-making process itself. These are the two main reasons for the course of such penetration processes.

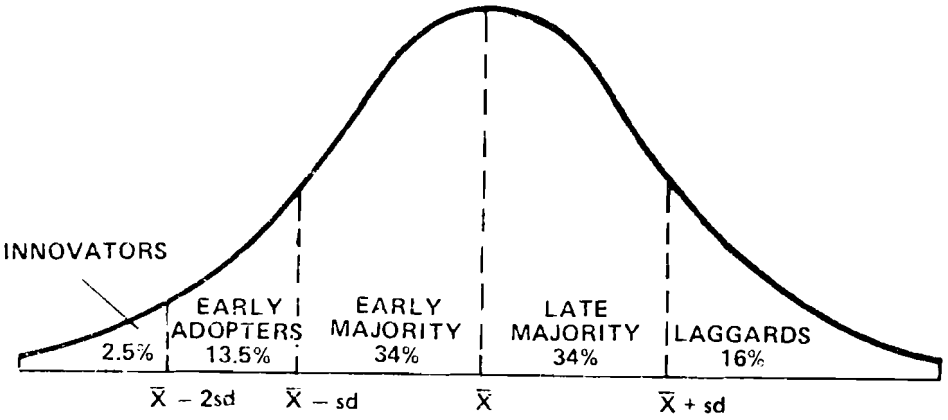
Of course, the behavior of individuals is not uniform. Casetti (1969) describes this behavior in the following way: (1) The adoption of an innovation by interested adopters results from messages of individuals who already have the innovation; (2) Potential adopters show a clearly differentiated resistance against the innovation; (3) Such a resistance can be observed everywhere; (4) This resistance can be overcome by frequent repetition of the messages.

After the individuals make their decision, the adopters may be subdivided into different groups. A possible and reasonable categorization was made by Rogers (1971). He assumes that a normal distribution for the stratum of adopters is sufficient, as measured by the time at which an individual adopts an innovation. The arrangement according to the single or double standard deviations from the mean value results in a categorization of adopters (Fig. 1).

This is, of course, an idealistic state compared with the real situation. The two groups showing an extreme behavior are the innovators and the laggards. The first 2.5% of adopters are innovators; the laggards on the other hand are those who can

only be made to adopt the innovation through pressure from their environment. An exact description of each group is given in Rogers' work (1971), pp. 183.

However, before an individual decides on whether to adopt or reject a certain innovation, the individual is influenced by the detailed information he has on the innovation. How can this process of information diffusion, which constitutes a special type of communication process, be described in a model?



The innovativeness dimension, as measured by the time at which an individual adopts an innovation or innovations, is continuous. However, this variable may be partitioned into five adopter categories by laying off standard deviations from the average time of adoption.

Figure 1. Adopter categorization on the basis of innovativeness. From Rogers (1971).

For a long time the two-stage model of the flow of communication was of great importance. According to this model the diffusion of information among potential adopters is the following: The mass media address the opinion leaders, i.e. people who have authority in their social environment and whose opinions are taken over by others; and the opinion leaders communicate their opinions through interpersonal channels. Several studies show, however, that this model is too simple to describe reality.

A model better suited to describe reality is the S.M.C.R.E. model, Rogers (1971). There the information on a certain innovation is passed on in the form of messages (M) from source (S) through channels (C) to receiver (R) where they cause certain effects (E), see Figure 2. The messages originating from the source (inventors, scientists, institutions, etc.) contain information on characteristic features, advantages, and disadvantages of the innovation in question.

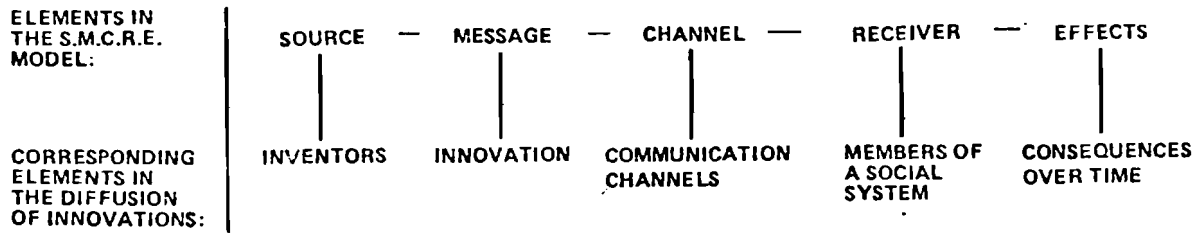
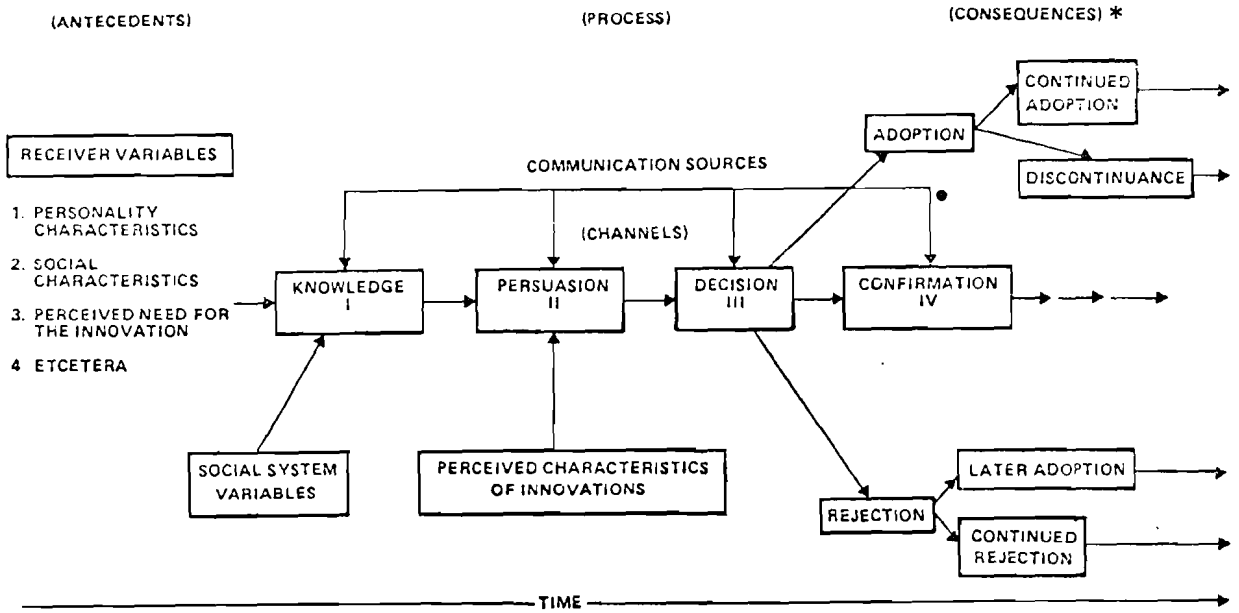


Figure 2. Elements in the diffusion of innovations and the S.M.C.R.E. communication model are similar. From Rogers (1971).

One can distinguish five main characteristics that are typical of all innovations: relative advantage, compatibility, complexity, testability, observability (for definition and comment see Rogers (1971), pp. 138). The mass media and interpersonal communication serve as communication channels. Mass media can reach big groups of potential adopters rapidly and serve as a general source of information accessible to everybody. With interpersonal contacts subjective influences are of special importance; reluctant and undecided individuals can be more easily persuaded. Thus it follows that opinion leaders or innovators can be addressed by mass media, while persons of the late majority or even in the group of laggards can, if at all, only be convinced of the advantages of an innovation by intensive personal contacts.

The effects observed in a receiver as a result of information may vary greatly. It depends on whether the receiver has already formed his opinion or not, i.e. on his stage of opinion-forming. For example, a person may be simply informed of the existence of a certain innovation, or there may be changes in a person's assessments and judgements, or the individual decides to adopt or reject an innovation. Thus the information reaching the individual through the diffusion process clearly serves as an input to his decision process, by which he processes the information on the economic and social environment and obtains as an output a decision to adopt or reject the innovation.

The course of the process is influenced by personal characterisitcs; its duration depends on the individual. Innovators, for example, pass this stage very fast, the process will not take long. The innovation decision process can be subdivided into various functions (stages). They are: knowledge, persuasion, decision, and confirmation (see Fig. 3).



*For the sake of simplicity we have not shown the consequences of the innovation in this paradigm but only the consequences of the process.

Figure 3. Paradigm of the innovation-decision process. From Rogers (1971).

During the first stage of knowledge the individual is informed that there is an innovation and collects the first information on it. The stage of persuasion is characterized by the adoption of a positive or negative attitude towards the innovation which is assessed on the basis of the main characteristics mentioned before. The next step towards adoption or rejection is the decision function. During the final stage of confirmation individuals check whether the right decision has been taken, which may either consolidate or change the opinion held.

It is important that the decision process can be interpreted as a learning process, i.e. a systematic change in the behavioral pattern. This systematic change is caused by stimuli the individual receives. During the decision process the stimuli correspond to the information given. According to learning theory the response triggered by the stimuli corresponds to the decision to adopt or reject an innovation.

Learning experiments have shown that the decision processes can be understood as learning processes. If a person faces a new task to which he or she is not accustomed, the person will make many mistakes in the beginning, but they will be fewer with increasing skill. The decrease in the number of mistakes (learning increments) is concerned and can be understood as being proportional: (a) to the skills already acquired; and (b) to the remaining skills still to be acquired.

Plotted versus time these learning increments lead to S-shaped curves, since the determined proportionality of learning increments corresponds to the assumptions made for the differential equation leading to the logistic function. It should be emphasized that this S-shaped course of learning processes was found in real learning situations.

How can the individual decision process be formally recorded? It is a process going on within the person deciding,

and is thus inaccessible for direct observation. Only the input (information) and the output (decision) of the process can be observed. In many such models, therefore, it is regarded as a black box which supplies an output that is only determined by the input. Thus the question is avoided as to how the input is processed and how a certain output comes about.

An exception are the buyer behavior models which try to describe the process in the black box. This is done in great detail in structure models (Howard/Sheth (1969, Nicosia (1966)). The disadvantage of these models is their deterministic and verbal character. Deterministic here means that the individual processes have to be accurately determined. Therefore, many of the postulated relations cannot be checked and are thus hypothetical. The verbal feature is a disadvantage because despite all attempts it has not been possible to come to an unambiguous clear-cut definition of terms. Furthermore, on account of the detailed data required it seems to be difficult to apply such models in practice.

Other than these deterministic models, the stochastic buyer behavior models avoid a detailed description of the black box processes. They are rather regarded as random processes, which is not to suggest that they are actually random processes. The data gathered to date do not warrnat such an assumption. The use of the stochastic approach actually indicates uncertainty with respect to the actual processes. Therefore, the output cannot be determined with certainty, but only with a certain probability. Another advantage of the stochastic approach is that one can confine oneself to the explicit determination of the most important relations, and all other influences are taken into account by the stochastic approach.

Possible stochastic formulations can be subdivided into Markovian models, learning models, and diffusion models. All three types of models have been used for commodities regularly consumed (coffee, beer, cigarettes, etc.) and not for durable

producsts which are important in this context. Markovian models can easily be handled mathematically, since the transition probabilities depend only on the most recent past.

With regard to commodities of regular consumption this is a severe restriction, since older changes in brands are not taken into account. This is no reservation on long-lived commodities since former purchases will hardly influence the present buyer decision. Technologies and environmental situations have also changed, and it is therefore difficult to compare the two buyer decisions.

The learning models are based on the stochastic theory of learning developed by Bush and Mosteller (1951, 1955); the best-known model is the linear learning model of Kuehn (1962) which includes the entire past, but it is more complicated mathematically. Furthermore, the model itself is not always consistent with the theory of Bush and Mosteller (see Topritzhofer (1974), pp. 164).

Diffusion models use another concept of learning theory, the stimulus sampling formulation by Estes (1963). The past is not taken into account, and there is no learning from experience. The individual has a number of buying arguments, and it is examined how the buying arguments are distributed over the competing brands. The product with most buying arguments is most likely to be bought. The most important studies in this field were compiled by Montgomery (1966, 1969) and Coleman (1964).

The practical applicability of these models is more restricted than that of the other models, mainly because it is difficult to determine the state space formed by the buying arguments. The following interpretation of the innovation decision process may be helpful to find a formulation that would correspond to the facts. This process can be understood as a process changing from a state of balance to imbalance and then to a new balance.

Before an innovation appears on the market, the potential adopter is in an equilibrium, i.e. the information (stimuli) received from the environment do not lead to any changes in the black box, they are consistent with the internal variable of the individual (attitudes, motivations, intentions, satisfaction, etc.). When the innovation appears, the potential adopter receives information causing changes of his internal variable. The system in the black box is no longer in an equilibrium. The more information is received, the bigger the imbalance becomes.

This situation changes when the individual adapts his internal state to the new situation by making a decision in favor or against the innovation. If the decision turns out to be right, the individual will again be in a state of equilibrium. In order to plot such a process in a formal model, one would have to use a time-dependent formulation of the relations to capture reality. This results in a time-dependent stochastic process which is consequently more difficult from the mathematical point of view. In the next section an attempt is made to describe the innovation decision process by means of the time-dependent Markovian process.

3. MATHEMATICAL FORMULATION

The following definitions are given by Ferschl (1970) and Opitz/Schader (1975):

Definition (a):

A stochastic process S is a set of random variables X_t with $S=\{X_t: t\in T\}$. $T\subset\mathbb{R}^1$ is the parameter space, whereby time t is frequently understood as a parameter. The value range Z of the random variables X_t for all $t\in T$ is called state space. The Markovian rpocesses are special stochastic processes.

Definition (b):

A stochastic process $S=\{X_t: t\in T\}$ is called Markovian process of the first order if with n=2,3,... it holds that for $t_1,\ldots t_n\in T$ with $t_1<t_2\ldots<t_n$

$$p(X_{t_n} \leq x | X_{t_{n-1}} = z_{n-1}, \ldots, X_{t_1} = z_1) = p(X_{t_n} \leq x | X_{t_{n-1}} = z_{n-1})$$

for any X and with $z_1 \ldots z_{n-1} \in Z$.

This means that the value of the random variable X_{t_n} only depends on the realization in the previous period but not on the realizations of periods that go back farther. This speciality is called Markovian property. It is, of course, also possible to define Markovian processes of the 0-th order (Bernoulli processes) and of higher order in this way.

Definition (c):

A stochastic process $S=\{X_t : t \in T\}$ is called Markovian process of the first order with a discrete time or Markovian chain of the first order, if it holds that $T = \mathbb{N} \cup \{0\}$, which means that $S=\{X_t : t=0,1,2,\ldots\}$, and if (b) is fulfilled, i.e. if it holds that

$$p(X_t = z_j | X_{t-1} = z_i, X_{t-2} = \ldots) = p(X_t = z_j | X_{t-1} = z_i) = p_{i,j}(t-1,t)$$

for all z_i, $z_j \in Z$.

The quantities $p_{i,j}(t-1,t)$ are called transition probabilities of the first order. Since they are time-dependent, they can be regarded as nonstationary transition probabilities. They can be summarized in a nonstationary transition matrix of the first order

$$P(t-1,t) = [(p_{i,j}(t-1,t)] \quad . \tag{1}$$

These nonstationary Markovian processes can be applied to various innovation decision processes. The innovations under consideration are durable ones that means innovations with long lifespans. The use of Markovian processes of the first order to describe decision processes involving products frequently purchased is not very realisitc, since the present purchase decision is influenced not only by the last purchase but also by all purchases during a certain period (year, month). But in the case

of durables earlier purchases are so far in the past that they have hardly any influence on the present decision. In many cases the old products and the new ones are so different that they cannot be compared. In these cases the deciding individual cannot apply his judgements derived from the use of the old product. The individual cannot apply any routinized decision behavior but rather has to show problem solving behavior (Howard/ Sleth, 1969). This means that the probability to adopt the innovation changes with the time. The conclusion from this is to apply a nonstationary Markovian process.

In the following the decisions are considered that can be characterized by two states. This situation includes all substitution processes where an innovation is substituted for an established product ("new" for "old"), and penetration processes where the state "old" means non-adoption and the state "new" means adoption of the innovation. In this context it should be emphasized that this formulation is not restricted to innovations in the economic sector but can also be applied to nonmaterial innovations (ideas, convictions, life styles, habits, etc.). Carson (1965) showed in impressive examples (penetration of modern math, programmed instruction) that these penetration processes also follow S-shaped curves. Furthermore, only those innovations are observed that proved successful on the market. This means in the substitution case that the old established product will be fully replaced by the innovation. In the case where no established product has to be substituted it means that all potential adopters will adopt the innovation sooner or later. For the nonstationary Markovian process it follows that an individual who has reached the state "new", i.e. the individual is in the possession of the innovation, is unlikely (with a probability of 1) to leave this state. Such a Markovian chain of a certain state which once reached cannot be left any more, is called absorbing Markovian chain. If the state "old" is expressed by zero and the state "new" by one we have the following situation (Fig. 4).

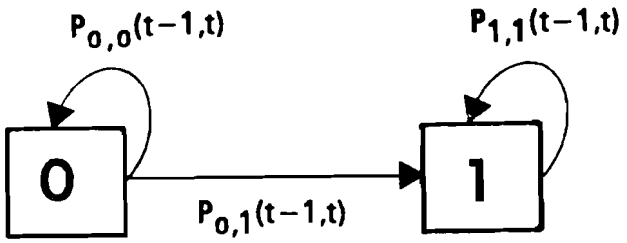


Figure 4.

Note that there is no arrow going back from state one to state zero. This expresses the absorbing ability of state one. The situation described in Figure 4 corresponds to the following transition matrix

$$P(t) = \begin{bmatrix} p_{0,0}(t-1,t) & p_{0,1}(t-1,t) \\ p_{1,0}(t-1,t) & p_{1,1}(t-1,t) \end{bmatrix} \qquad (2)$$

Since the Markovian chain is absorbing ($p_{1,1}(t-1,t)=1$) and $p(t)$ is a stochastic matrix (the sums of rows equal one), the process is controlled by the transition probability from state zero to one. It is assumed that this probability has the following form:

$$p_{0,1}(t-1,t) = \alpha f_1(t-1)+\beta \ , \qquad (3)$$

where

α,β are parameters, and
$f_1(t-1)$ is the market share of the state "new" in period $(t-1)$.

Equation (3) refers to the innovation decision process analyzed above. This leads to the following interpretation of the two parameters α and β. If the market penetration of an innovation is still small, the transition probability is determined by parameter β. Thus β can be interpreted as a measure for the individual's innovativeness. This causes an increased receptivity for stimuli from the environment, which results in increased interest (Howard/Sleth (1969), p. 286). One of the

results of many investigations has been that novelty is an important reason for many consumers to adopt an innovation. Especially in the beginning when an innovation enters the market it is the novelty that causes the first purchases because the individual is not yet in a position to obtain and process the necessary information to come into objective judgements about the innovation. As already mentioned, only a few individuals will purchase an innovation in such an early stage. Sometimes you can realize that the effect of novelty is reinforced by an effect of boredom caused by dissatisfaction with the established products. Parameter β can also be understood as a measure for the individual's propensity to take risks. This propensity depends on the social and economic importance of the innovation. If the innovation possesses great social and economic impacts, the individual will try to minimize the risks. In such a case people generally show an aversion towards risk. Nevertheless some are more likely to take risks than others (comparison of innovators with laggards). So parameter β can help characterize the potential adopter. A high value of β signals a high propensity to take risks.

The other parameter α should be seen in connection with the market share reached in the previous period. As the penetration process increases, the contribution of equ.(3) towards transition probability also increases. Parameter α thus expresses the increased effect of information (stimuli) from the environment. These stimuli reach the individual mainly by interpersonal communication with people who have already made their decision to adopt or reject the innovation. The degree of penetration of the innovation is expressed by the market share of the previous period. As parameter α is linked multiplicatively with the market share of the previous period, α can be interpreted as a measure of the individual's propensity to imitate other people. This imitation behavior has to be understood as a learning process. Thus one important aspect of the learning process which takes place during an innovation decision in reality is projected in the formal model of a nonstationary Markovian process explicitly. A high value of α then signals a high propensity to

imitate. Both parameters α and β are constants. It is assumed that they do not vary during the decision process on the innovation. The character of this process as a learning process is implied by the mathematical formulation. With the Markovian chain, which means that the transition probability to a certain state depends on the previous state, the effect of learning by experience is taken into account. The matrix $P(t)$, see equ.(2), can be written as follows:

$$P(t) = \begin{bmatrix} 1-\alpha f(t-1)-\beta & \alpha f(t-1)+\beta \\ 0 & 1 \end{bmatrix} . \qquad (4)$$

The temporal development of this nonstationary Markovian chain cannot be defined in mathematical terms as explicitly as can be done if stationary Markovian processes are involved. But it is possible to obtain some information on the behavior in the long run if the changes of the transition matrix (4) are considered from period to period. These changes can be understood by applying the concept of causative matrices (see Lipstein (1965, 1968), Hurary, Lipstein, Styan (1970)). The causative matrix $C(t)$, which describes the change in the transition matrix (4) from period t to period t+1, is obtained by

$$\begin{aligned} P(t)\ C(t) &= P(t+1) \quad ; \\ C(t) &= P(t)^{-1} P(t+1) \quad . \end{aligned} \qquad (5)$$

From (5) it is obvious that in the case of stationary transition matrixes the causative matrix $C(t)$ corresponds to the identity matrix. Conceptually the causative matrices can be compared with the derivatives of a function. They have row sums of one, but they are not necessarily stochastic matrices, since the inverse matrix of a nonsingular stochastic matrix may also have negative elements (see Lipstein (1968)). In order to obtain the causative matrices (5), the inverse of each transition matrix must exist, i.e. the transition matrix must be nonsingular. In the specific case of (4) nonsingularity is always safeguarded

except if the transition probability from state zero to state one ($p_{0,1}(t-1,t)$ becomes one. However, in this case the causative matrix is meaningless, for a stable state has been reached, i.e. state one ("new") has been reached with certainty. For general cases McKenzie (1960) gives the following criterion which guarantees nonsingularity: If matrix A possesses a dominant diagonal (the elements of the main diagonal are greater than the row sums of the remaining elements in the respective rows), A is nonsingular (possesses an inverse). According to (5) one obtains:

$$C(t) = \begin{bmatrix} \frac{1}{1-\alpha f_1(t-1)-\beta} & -\frac{\alpha f_1(t-1)+\beta}{1-\alpha f_1(t-1)-\beta} \\ 0 & 1 \end{bmatrix}$$

$$* \begin{bmatrix} 1-\alpha f_1(t)-\beta, & \alpha f_1(t)+\beta \\ 0 & 1 \end{bmatrix} ;$$

$$C(t) = \begin{bmatrix} \frac{1-\alpha f_1(t)-\beta}{1-\alpha f_1(t-1)-\beta} & \frac{\alpha(f_1(t)-f_1(t-1))}{1-\alpha f_1(t-1)-\beta} \\ 0 & 1 \end{bmatrix} . \qquad (6)$$

It follows from (6) that C(t) is a stochastic matrix where $f_1(t)$ is bigger than $f_1(t-1)$. The nonstationary Markovian chain generates a sequence of causative matrices. Their long-term behavior is characterized by the eigenvalues of C(t) (Lanczos (1956), McKenzie (1960)). The eigenvalues are the solutions of the characteristic polynomial. In this case the following characteristic polynomial is obtained:

$$
\begin{aligned}
|C(t)-\lambda(t)I| &= 0 \\
&= \lambda(t)^2 - \left(\frac{1-\alpha f_1(t)-\beta}{1-\alpha f_1(t-1)-\beta} + 1\right) \lambda(t) \\
&\quad + \frac{1-\alpha f_1(t)-\beta}{1-\alpha f_1(t-1)-\beta} = 0
\end{aligned} \tag{7}
$$

where

$\lambda(t)$ are eigenvalues and

I the identity matrix.

The solutions of (7) are

$$
\lambda_1(t) = 1, \quad \lambda_2(t) = \frac{1-\alpha f_1(t)-\beta}{1-\alpha f_1(t-1)-\beta} \quad . \tag{8}
$$

From (8) we draw the following conclusion:

$$
\begin{aligned}
&\text{(a)} \quad f_1(t) = f_1(t-1) \nearrow\searrow \lambda_2(t) = 1 \quad ; \\
&\text{(b)} \quad f_1(t) > f_1(t-1) \nearrow\searrow \lambda_2(t) < 1 \quad ; \\
&\text{(c)} \quad f_1(t) < f_1(t-1) \nearrow\searrow \lambda_2(t) > 1 \quad .
\end{aligned}
$$

Case (a) characterizes a stable situation. The market share of the innovation does not change from period to period. Case (c) characterizes a situation where C(t) is no longer stochastic because the element (1.2) of C(t) is negative. C(t) converges (see Lipstein (1968)) if the absolute value of each of its eigenvalues is smaller than one or equals one. This is case (b). It follows that in this case only successful innovations are considered because the market share has to be higher or at least as high as in the previous period. This implies that with an increasing market share of the innovation the element (1.1) of the matrix P(t) (see equ.(4)) decreases. This means that the individual's loyalty to state zero (the established product) decreases and the readiness to change to state one (decision in favor of the innovation) increases. It follows that the individual has adopted the innovation with probability one at the end of the penetration process. Then the nonstationary Markovian process has developed into a Bernoulli process (Markovian process of zeroth order).

By means of the nonstationary Markovian chain the penetration process can be characterized as a process leading from a state of equilibrium via a state of disequilibrium to a new state of equilibrium. To illustrate this point of view the individual's state vector at any period t should be considered. This state vector w(t) expresses the probability that the individual is in one of the given states. At the time t=0, the adequate state vector w(0) is called initial vector. For the innovation decision process this initial vector w(0) = (1,0), i.e. the individual is with certainty in the state zero. This is reasonable, since at time t=0 the innovation has not yet entered the market. It is impossible for the individual to adopt the innovation. This initial state w(0) = (1,0) also characterizes a state of equilibrium. The individual is in a stable state because there are no stimuli that do not agree with its judgements about the established product or situation. As soon as the innovation appears on the market, informations about the innovation reach the individual and cause conflicts in the individual's pattern of behavior. The individual feels doubts whether its judgements about state zero are still justified. The transition mechanism starts and the individual is no longer in a state of equilibrium. This is expressed by the eigenvalue $\lambda_2(t)$ of C(t), which in the case of a successful innovation is smaller than one. The more $\lambda_2(t)$ deviates from one, the bigger is the state of disequilibrium caused by the penetration of the innovation. This imbalance is also expressed by the state vector which changes from period to period by application of the transition matrix (4) to the state vector of the previous period. Thus w(t) can be determined from

$$w(t) = w(t-1)\ P(t) \quad . \tag{9}$$

When most of the people have adopted the innovation, the market share increments are getting smaller and smaller, i.e. the market share remains more or less constant. This is also expressed by the eigenvalue $\lambda_2(t)$, which approaches one. This means that the

individual reaches a new state of equilibrium. This state is characterized by the state vector w(t) at the time of saturation. But there exists only a state of equilibrium if the individual is satisfied with his decision, i.e. he does not feel any cognitive dissonance. If this is the case, one obtains the state vector $w(t_{saturation}) = (0,1)$, i.e. the individual is with probability one in state one. It should be emphasized that all potential adopters may be characterized by a state vector at the time t, with the transition mechanism being projected by means of matrix P(t).

If this nonstationary Markovian chain is to be applied to real penetration processes, description of the development of aggregated data like market shares or sales figures is necessary. Therefore it is necessary to aggregate the individual approach. With N potential adopters, the aggregation occurs via N parallel processes characterized by N transition matrices P(t). If one assumes homogeneity and independence of the individuals, all N potential adopters can be characterized by the same transition matrix P(t). Applying this transition matrix to the number of people in states zero and one, the average number of people who are in one of the two states at time t is obtained. By dividing these numbers of individuals by the total number of potential adopters, fractions of the total number are obtained. If it is assumed that each of the potential N adopters can adopt only one issue of the innovation so that N is the saturation limit of the innovation, the fraction can be understood as a market share. If an individual purchases ten issues of an innovation, it is replaced by ten identical individuals. So it is possible to obtain instead of the state vector of a single individual the vector of expected market shares. *The constant parameters* α *and* β *are estimated from a time series of aggregated data (market shares)* as will be shown below. They are equal for each of the N potential adopters and can be considered average values. They characterize an average adopter. It should be emphasized that the assumptions of homogeneity and independence constitute a strong limitation of reality. To be more realistic would mean

to assume heterogeneity of the parameters α and β. In this case α and β would follow a certain probability distribution (e.g. the beta-distribution) but then the nonstationary Markovian chain would become too complex with respect to its applicability. Whether the chosen aggregated formulation has thus become too unrealistic will be shown by way of concrete examples.

Instead of the individual state vectors w(t) we consider market share vectors in the aggregated version as already indicated. The market share vectors f(t) are obtained according to (9). The initial vector f(0) = (1,0) indicates that the state zero has a market share of 100%, i.e. that all individuals are with certainty in state zero. This has to be understood in the aggregated version too as an expression of a stable situation. When the innovation has not yet entered the market there is a state of equilibrium. More or less heavy disturbances of the market are caused by the appearance of the innovation. Market shares are no longer stable, they change from period to period. If the innovation is successful it will increase its market share continuously. This state of disequilibrium of the market is characterized by the corresponding market share vectors. When the increase of the market share of the innovation decreases continuously, the limit of saturation is approached. The market share vectors begin to remain constant which means that a state of a new market equilibrium is reached. This is expressed by the market share vector at the time of saturation $f(t_{sat.})=(0,1)$, i.e. the successful innovation has reached a market share of 100%. It should be mentioned that there exists a special index, the rate of convergence (Topitzhofer, 1964), to describe the behavior of convergence from the state of disequilibrium to the state of equilibrium. This index can be applied to the causative matrix C(t) of equ.(6) if C(t) is stochastic. It holds for the eigenvalues of a stochastic matrix that their absolute values are less or equal one. Hence the rate of convergence follows:

$$\prod_{i=1}^{n} |\lambda_i| \leq 1 .$$

The conclusion is that the more the rate of convergence approaches one, the slower the state of equilibrium will be reached (Topitzhofer, 1973, p.153).

To describe the penetration process of an innovation you can determine the market share of the innovation at any time t. According to (9) we obtain:

$$f(t) = f(t-1)\ P(t) \tag{10a}$$

or

$$(f_o(t), f_1(t)) = (f_o(t-1), f_1(t-1)) \begin{bmatrix} 1-\alpha f_1(t-1)-\beta, \alpha f_1(t-1)+\beta \\ 0 \qquad\qquad 1 \end{bmatrix}. \tag{10b}$$

The market share of the innovation in period t, $f_1(t)$ is:

$$f_1(t) = f_o(t-1)\ (\alpha f_1(t-1)+\beta+f_1(t-1) \tag{11}$$

Only the market share of the innovation has to be considered because the market share of the state zero, i.e. the established situation or product, at time t, $f_o(t)$ results from the difference between 1 (the limit of saturation) and the market share of the innovation $f_1(t)$. It becomes evident from (11) that $f_1(t)$ constitutes a mean value (the expected market share), as $f_1(t)$ results from multiplication of the market shares of the previous period by the corresponding probabilities. To gather information about the temporal development of $f_1(t)$ we consider the increase of $f_1(t)$ from period (t-1) to period (t) and obtain:

$$\begin{aligned} f_1(t) - f_1(t-1) &= f_o(t-1)\ (\alpha f_1(t-1)+\beta) \\ &= \beta(1-f_1(t-1)+\alpha f_1(t-1)(1-f_1(t-1)) \quad . \end{aligned} \tag{12}$$

If the values for the parameters α and β are known equ. (12) describes the temporal development of $f_1(t)$. We will obtain a curve from which we want to find its properties (derivatives, maximum, point of inflection). We therefore consider the continuous quotient instead of the difference quotient, and change

from the difference equation (12) to the following differential equation:

$$\dot{f}_1(t) = \frac{df_1(t)}{dt} = \alpha f_1(t)(1-f_1(t)) + \beta(1-f_1(t)) \quad . \tag{13}$$

With (13) we obtain the market share rate of the innovation at time t. This differential equation has the following solution (see Bass (1969), Opitz/Schader (1975)):

$$f_1(t) = \frac{1-\exp(-(\alpha+\beta)t)}{\frac{\beta}{\alpha}\exp(-(\alpha+\beta)t)+1} \quad . \tag{14}$$

In order to find the point of inflection of (14) we form the derivatives:

$$\dot{f}_1(t) = \frac{(\alpha+\beta)^2 \exp(-(\alpha+\beta)t)}{\frac{\alpha}{\beta}\exp(-(\alpha+\beta)t)+1)^2} \quad . \tag{15}$$

(15) yields another expression of the market share rate. It can be compared with (13). Differentiating (15) we obtain:

$$\frac{d(\dot{f}_1(t))}{dt} = \frac{\frac{(\alpha+\beta)^3}{\beta}\exp(-(\alpha+\beta)t)(\frac{\alpha}{\beta}\exp(-(\alpha+\beta)t)-1)}{(\frac{\alpha}{\beta}\exp(-(\alpha+\beta)t)+1)^3} \quad . \tag{16}$$

Setting (16) equal to zero we obtain the point of inflection:

$$\frac{d(\dot{f}_1(t))}{dt} = 0 \qquad t^* = \frac{1}{\alpha+\beta}\ln\frac{\alpha}{\beta} \quad . \tag{17}$$

It follows from (17) that t* becomes greater than zero, if α is greater than β.

Substituting (17) for (15) we obtain the maximum market share rate:

$$\dot{f}_1(t^*) = \frac{(\alpha+\beta)^2}{4\alpha} \quad . \tag{18}$$

Substituting (17) for (14) we obtain the market share at the point of inflection. The value obtained at the point of inflection is important because it can be compared with the corresponding values of other growth functions, especially the logistic function. The following market share is obtained at the point of inflection:

$$f_1(t^*) = \frac{\alpha-\beta}{2\alpha} \quad . \tag{19}$$

It follows from (19) that analogous to the logistic function we obtain a value of the market share of 50% at the point of inflection, if β is very small in relation to α. This means that the proportion of α and β determines whether the value of the market share at the point of inflection approaches the limit of 50%.

$$\alpha >> \beta \nearrow f_1(t^*) = \frac{\alpha-\beta}{2\alpha} \approx \frac{1}{2} \quad . \tag{20}$$

Before we can apply the stochastic process characterized by (10) to practical cases in order to test the validity of the non-stationary Markovian chain, we have to estimate the two parameters α and β. The parameters can be estimated by applying the principle of ordinary least squares. For this purpose (12) is written as follows:

$$\frac{f_1(t)-f_1(t-1)}{1-f_1(t-1)} = \alpha f_1(t-1)+\beta \quad . \tag{21}$$

Setting

$$\frac{f_1(t)-f_1(t-1)}{1-f_1(t-1)} = \hat{Y}_t \hat{Y}_t$$

$$\alpha = a$$

$$\beta = b$$

$$f_1(t-1) = X_t X_t \quad ,$$

we obtain the following straight line:

$$\hat{y}_t = ax_t + b \quad . \tag{22}$$

The distances between the values calculated by means of (21) and the actual values Y_t have to be minimized. We obtain:

$$\min_{a,b} \sum_{y_t} (y_t - ax_t - b)^2 = :\min Q \tag{23}$$

The partial derivatives of (23) result in a system of normal equations. From this we obtain the following estimated values for α and β:

$$a = \frac{\sum_t x_t y_t - \bar{T}\bar{x}\bar{y}}{\sum_t x_t^2 - \bar{T}(\bar{x})^2}$$

$$b = \bar{y} - a\bar{x} \tag{24}$$

with

$$\bar{x} = \frac{1}{\bar{T}} \sum_t x_t$$

$$\bar{y} = \frac{1}{\bar{T}} \sum_t y_t$$

$\bar{T}$ = horizon of observation (number of time periods).

With (24) we obtain estimates for the parameters α and β, and the market shares of the innovation at the time t can be calculated according to (11). The goodness of fit between the calculated course and the actual course is examined by means of the chi-square-test. The chi-square statistic is defined as follows:

$$\chi_o^2 = \sum_{t=1}^{\bar{T}} \frac{(f_1(t)_{\text{actual}} - f_1(t)_{\text{calculated}})^2}{f_1(t)_{\text{calculated}}} \quad . \tag{25}$$

On the basis of the zero hypothesis that the nonstationary Markovian chain described by (10), (11) is consistent with the

data given, it is examined whether the value χ^2_o calculated according to (25) is higher than the bound ($\chi^2_{.05}$ or $\chi^2_{.01}$) belonging to a significant value λ (5%, 1%) to be determined. If this is the case one must abandon the zero hypothesis that the model is consistent with the data with a probability of $1-\lambda$ (95%, 99%). Now we are able to investigate some practical cases in order to test the validity of the model.

4. APPLICATIONS

We applied the model to five practical examples. These are the substitution processes of material innovations, and the diffusion of two nonmaterial innovations. The data for these studies are from the United States, because these empirical publications were readily available.

First case:

The substitution of synthetic fibres for natural fibres in the US between 1930 and 1965 is described. The data are taken from the study by Fisher and Pry (1975) mentioned above (see Table 1 below). They found that the logistic function fits the data very well. The nonstationary Markovian chain produces a χ^2_o-value which is clearly below the bound of the significance level of 5 percent:

$$\chi^2_o = 6{,}366 < \chi^2_{0.05} = 11{,}07, \quad 5 \text{ degrees of freedom} \quad .$$

It follows that the stochastic model is consistent with the data (see Figure 5). The quotient of the two parameters α and β is:

$$\begin{aligned} \alpha &= 0.3479644 \\ \beta &= 0.0257398 \end{aligned} \qquad \frac{\beta}{\alpha} = 0.074 \quad .$$

This means that the value of β is 7.4% of the value of α. It follows for the market share of synthetic fibres at the point of inflection t*:

$$f_{\text{fibre}_{\text{synth}}}(t^*) = 46.3\% \quad ,$$

i.e. the point of inflection is reached somewhat earlier than in the case of the logistic function.

Table 1. Market shares of synthetic fibres.

year	fibre synth. actual	fibre synth. calculated
1930	4.4%	5.18%
1935	7.9%	9.99%
1940	10%	14.8%
1945	14%	19.42%
1950	22%	23.85%
1955	28%	28.09%
1960	29%	32.15%
1965	43%	36.02%

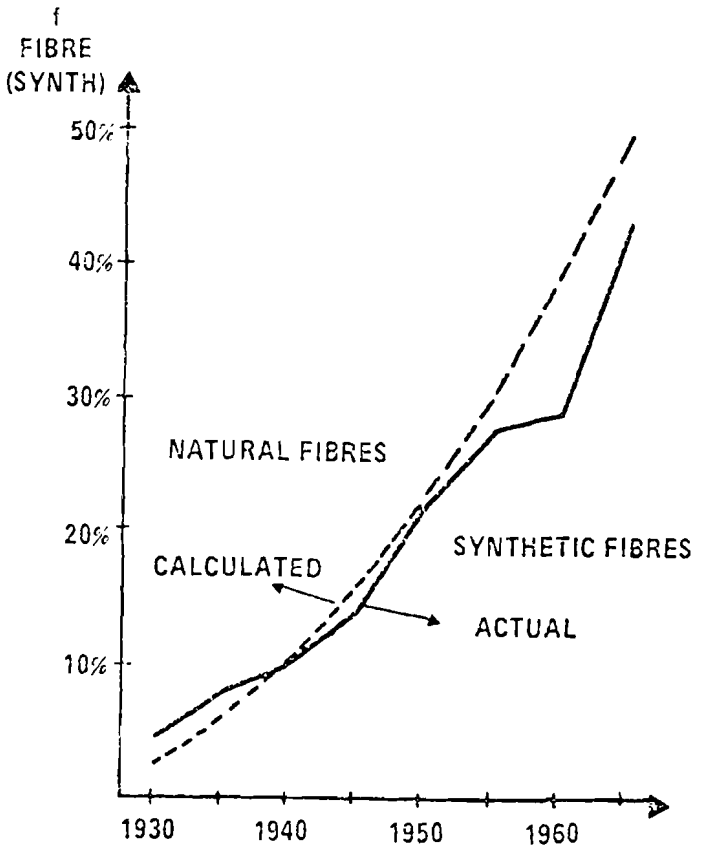


Figure 5. Substitution of synthetic fibres for natural fibres.

Second case:

The substitution of stem ships for sailing ships in the US between 1800-1960 is considered. The data are taken from US Department of Commerce (1975), see Table 2. Here too the result obtained shows the consistence of the stochastic model with the given data:

$$\chi_o^2 \; 0 \; 6.10 < \chi^2_{0.05} = 22.36, \quad 13 \text{ degrees of freedom.}$$

As this process has reached the limit of saturation in reality, the stochastic process can be applied to the whole penetration process of the innovation (steam ships) from entering the market until saturation. The curve is as expected S-shaped, and the value of market share at the point of inflection t* approaches nearly the 50% value of the logistic function. The value obtained at the point of inflection is:

$$f_{ship_{steam}}(t^*) = 48.88\%$$

This is expressed also by the quotient of α and β, which is remarkably lower than the quotient in the first case. The following quotient is obtained:

$$\begin{aligned} \alpha &= 0.4974329 \\ \beta &= 0.011186 \end{aligned} \qquad \frac{\beta}{\alpha} = 0.225 \quad ,$$

i.e. the value of β is 2.25% of the value of α.

Table 2. Substitution of steam ships for sailing ships.

year	$f_{ship_{steam}}$ actual	$f_{ship_{steam}}$ calculated	year	$f_{ship_{steam}}$ actual	$f_{ship_{steam}}$ calculated
1810	0.1%	1,12%	1900	58.5%	66.29%
1820	1.7%	2.78%	1910	75.5%	77.78%
1830	5.4%	5.21%	1920	91.0%	86.62%
1840	10.2%	8.73%	1930	94.5%	92.53%
1850	14.9%	13.71%	1940	99.5%	96.05%
1860	16.2%	20.56%	1950	99.7%	97.98%
1870	31.2%	29.57%	1960	99.9%	98.99%
1880	34.0%	40.72%			
1890	46.9%	53.39%			

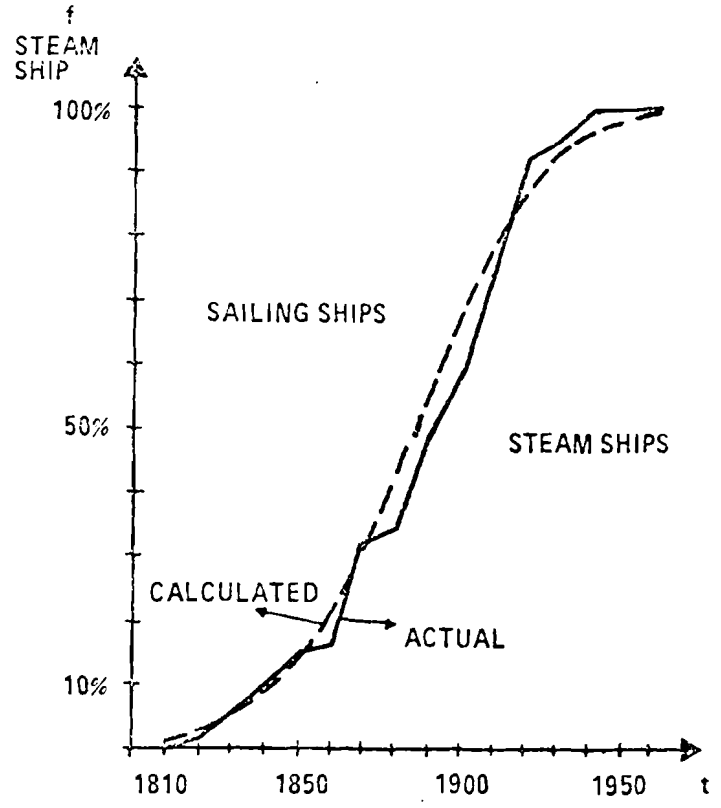


Figure 6. Substitution of steam ships for sailing ships.

Third case:

This example does not describe a substitution process but the spreading of an innovation which did not directly displace an established product but rather filled a gap. The penetration process of the telephone in the US is examined for the period from 1880 to 1965 (see Figure 7). The data are taken from Martino (1968), see Table 3. The percentages of telephones given there refer to 1000 people. The saturation limit is assumed at a value of 696.9 telephones per 1000 people. The goodness-of-fit test again shows the consistence with the actual values:

$$\chi_0^2 = 20.45 < \chi_{0.05}^2 = 25.15 \text{ degrees of freedom.}$$

The value of β is 5.4 percent of the value for α with

$$\alpha = 0.1860416 \quad \text{and}$$
$$\beta = 0.0100194 \quad .$$

This leads to the presumption that the value of market share at the point of inflection t* will be somewhat lower than 50%. Actually the following values are obtained:

$$f_{\text{telephone}}(t^*) = 47.3\% \quad .$$

Table 3. Market penetration of telephone.

year	$f_{telepone}$ actual	$f_{telephone}$ calculated	year	$f_{telephone}$ actual	$f_{telephone}$ calculated
1880	0.16%	1.0%	1930	23.3%	24.8%
1885	0.28%	2.1%	1935	19.6%	29.2%
1890	0.5%	3.52%	1940	23.7%	33.7%
1895	0.7%	5.05%	1945	28.4%	38.3%
1900	2.5%	7.04%	1960	40.3%	43.4%
1905	7.0%	9.2%	1955	48.4%	48.4%
1910	11.8%	11.85%	1960	58.5%	53.5%
1915	14.9%	14.8%	1965	68.6%	58.8%
1920	17.7%	17.8%			
1925	20.8%	20.75%			

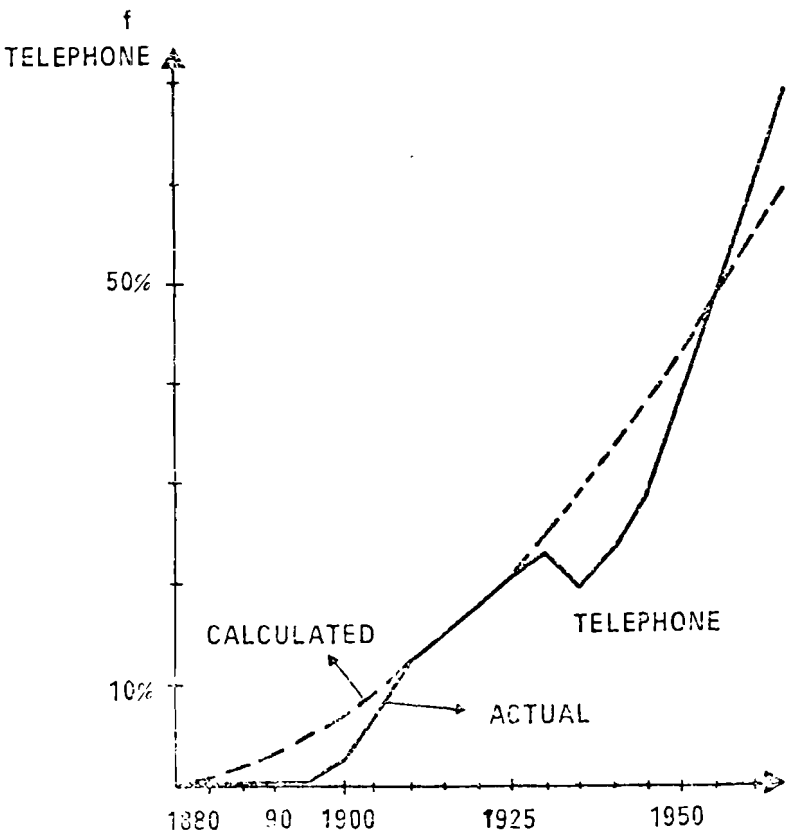


Figure 7. Market penetration of telephone.

Fourth case:

Here the spreading of a nonmaterial innovation is considered. It is the adoption of modern math in American high schools and colleges (see Figure 8). This case is reported by Carlson (1965), see Table 4. Within the period from 1958 to 1963 the degree of penetration reached 74.5 percent of the schools under consideration. This rapid penetration of educational methods at this time can be explained by the start of Sputnik I. The American society was shocked and all possible efforts were undertaken to fill the gap in science and technology. In the early sixties these efforts seem to decrease. After annual increases between 15 and 25 percent from 1959 to 1962 in 1963 only 9 percent increase is observed. This is the reason why the goodness-of-fit is only tested until 1962. If data after 1963 were available the model would be able to include this smoothing effect. For the time period from 1958 to 1962 the model is consistent with the data:

$$\chi_0^2 = 7.33 < \chi_{0.05}^2 = 12.59, \quad 6 \text{ degrees of freedom.}$$

The value of β is very small in relation to α (2.14% of the value for α) with

$$\alpha = 0.6666443 \quad \text{and}$$
$$\beta = 0.0142879 \quad .$$

Thus it follows that the degree of penetration at the point of inflection t* is:

$$f_{\text{modern math}}(t^*) = 48.93\% \quad .$$

Table 4. Penetration of modern math.

year	$f_{mod.math}$ actual	$f_{mod.\ math}$ calculated	year	$f_{mod.math}$ actual	$f_{mod.\ math}$ calculated
1958	1.4%	1.42%	July 1960	32.24%	35.24%
July 1958	3.45%	3.75%	1961	45.61%	51.38%
1959	6.54%	7.53%	July 1961	56.45%	68.73%
			1962	65.23%	83.5%
July 1959	13.0%	13.49%	July 1962	70.28%	-
1960	21.31%	22.51%	1963	74.5%	-

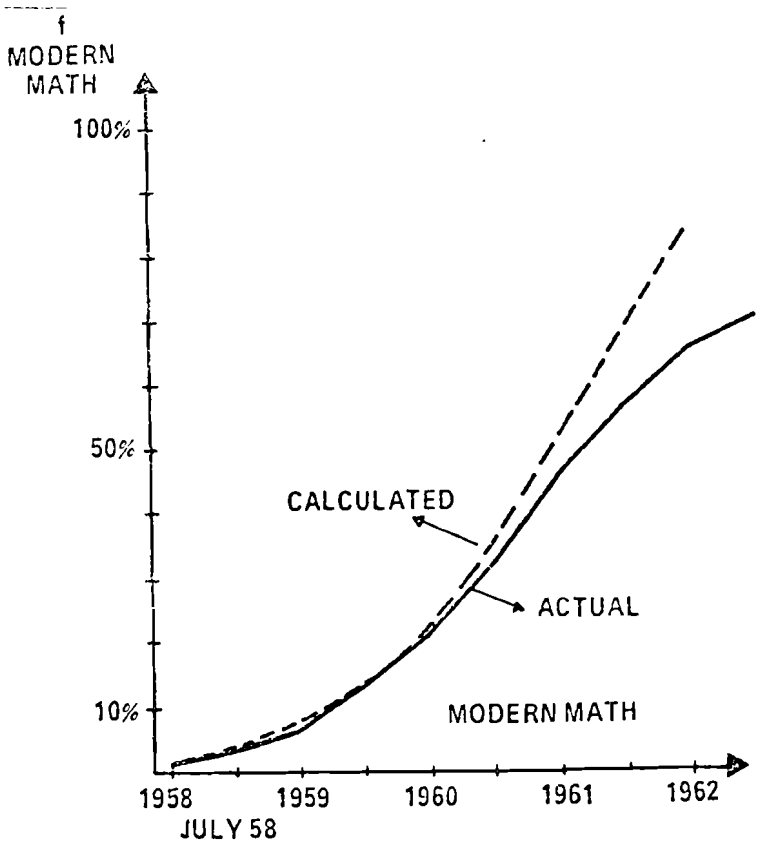


Figure 8. Penetration of modern math.

Fifth case:

Here another penetration process of an educational innovation is described. It is the spreading of programmed instructions from 1958 to 1963 in the US (see Figure 9) reported again by Carlson (1965). The rate of diffusion is not so fast compared with the spreading of modern math. But again one realizes after 1962 a smoothing effect. Probably the same arguments hold for the temporal development of this innovation as in the case of modern math. So again the data from 1958 to 1962 are applied to test the goodness-to-fit (see Table 5). The model is again consistent with the data:

$$\chi_0^2 = 3.67 < \chi_{0.05}^2 = 12.59, \quad 6 \text{ degrees of freedom.}$$

The value of β is 1.41% of the value of α with

$$\alpha = 0.5902603 \quad \text{and}$$
$$\beta = 0.0083161 \quad .$$

This implies a value at the point of inflection t* of nearly 50%. It is exactly:

$$f_{\text{progr.instr.}}(t^*) = 49.3\% \quad .$$

Table 5. Penetration of programmed instructions.

year	$f_{progr.instr.}$ actual	$f_{progr.instr.}$ calculated	year	$f_{progr.instr.}$ actual	$f_{progr.instr.}$ calculated
1958	2.05%	0.83%	July 1961	36.26%	39.96%
July 1958	2.34%	2.14%	1962	44.58%	52.18%
1959	3.55%	4.18%	July 1962	48.04%	-
July 1959	7.01%	7.34%	1963	49.53%	-
1960	12.62%	12.16%			
July 1960	18.97%	19.19%			
1961	26.17%	29.01%			

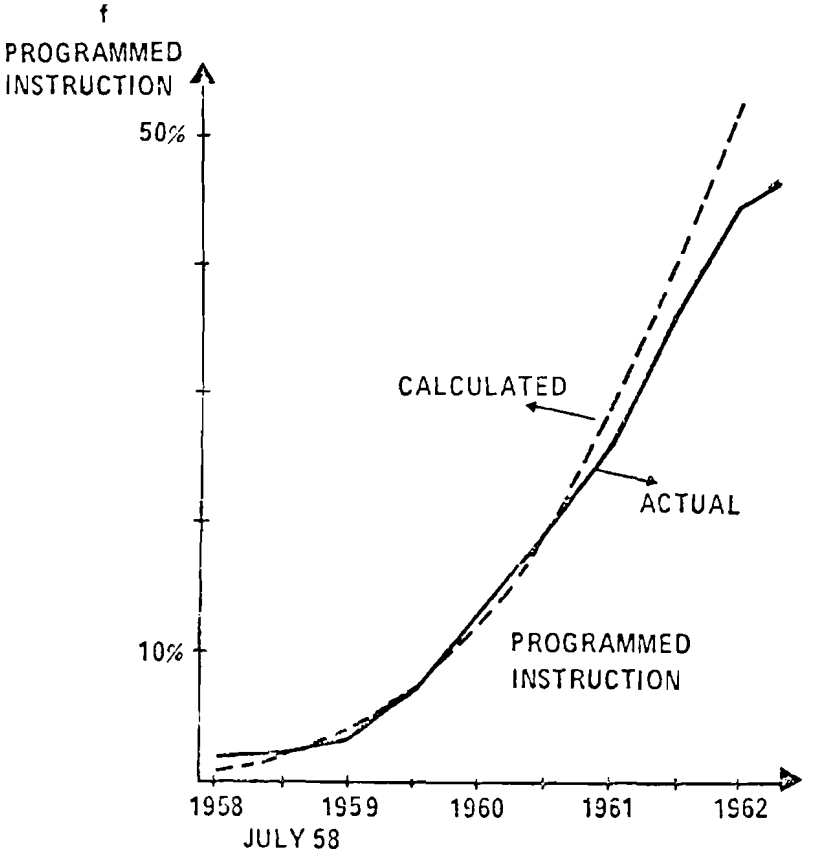


Figure 9. Penetration of programmed instructions.

LIST OF SYMBOLS

S = stochastic process

X_t = random variable at time t

T = parameter space

Z = state space

I = identity matrix

$P(t)$ = transition matrix at time t

$C(t)$ = causative matrix at time t

α, β = parameters

t = time

z = elements of the state space

$p(t)$ = probability at time t

$f(t)$ = vector of market shares at time t

$f_o(t)$ = market share of state zero at time t

$f_1(t)$ = market share of state one at time t

$w(t)$ = state vector of the individual at time t

$\lambda(t)$ = eigenvalue vector at time t

REFERENCES

Bailey, N.T.J. (1950), *A Simple Stochastic Epidemic*, Biometrika 37.

Bartlett, M.S. (1946), *Stochastic Processes*, University of North Carolina.

Bass, F.M. (1969), *A New Product Growth Model for Consumer Durables*, Management Science 15.

Blackman, A.W. (1974), *The Market Dynamics of Technological Substitution*, Technological Forecasting and Social Change 6.

Bush, R.R. (1951), *A Mathematical Model for Simple Learning*, Psychological Review 58.

Carlson, R.O. (1965), *Adoption of Educational Innovations*, Eugene, University of Oregon.

Casetti, E. (1969), *Why do Diffusion Processes Conform to Logistic Trends?* Geogr. Anal. 1.

Coleman, J.S. (1964), *Introduction to Mathematical Sociology*, New York.

Estes, W.K. (1963), *Stimulus Sampling Theory*, Handbook of Mathematical Psychology, R.D. Luce, R.R. Bush, E. Galanter (eds.) Vol. II.

Ferschl, F. (1970), *Markovketten*, Lecture Notes in OR and Mathematical Systems, 35.

Fisher, J.C., and R.H. Pry (1970), *A Simple Substitution Model of Technological Change*, Technological Forecasting and Social Change 3.

Griliches, Z. (1957), *Hybrid Corn: An Exploration in the Economics of Technological Change*, Econometrica 25.

Hägerstrand, T. (1952), *The Propagation of Innovation Waves*, Lund Studies in Geography, Series B, No.4.

Hägerstrand, T. (1967), *Innovation Diffusion as a Spatial Process*, Chicago.

Harary, F., B. Lipstein and G.B.H. Styan (1970), *A Matrix Approach to Nonstationary Chains*, Operations Research 18.

Haskey, H.W. (1954), *A General Expression for the Mean in a Simple Stochastic Epidemic*, Biometrika 41.

Howard, J.A. and J.N. Sheth (1969), *The Theory of Buyer Behavior*, New York.

Hudson, J.C. (1969), *Diffusion in a Central Place System*, Geogr. Analysis 1.

Kuehn, A.A. (1962), *Consumer Brand Choice - A Learning Process?* Journal of Advertising Research, Vol. 2.

Lanczos, C. (1956), *Applied Analysis*, Prentice Hall.

Listein, B. (1965), *A Mathematical Model of Consumer Behavior*, Journal of Marketing Research, Vol. 2.

Listein, B. (1968), *Test Marketing: A Perturbation in the Market Place*, Management Science.

Mansfield, E. (1961), *Technical Change and the Rate of Imitation*, Econometrica 29.

Martino, J.P. (1968), *Technological Forecasting for Decision Making*.

McKenzie, L. (1960), *Matrices with Dominant Diagonals and Economic Theory*, Mathematical Methods in the Social Sciences, Stanford University Press.

Montgomery, D.B. (1966), *A Probabilistic Diffusion Model of Dynamic Market Behavior*, MIT.

Montgomery, D.B. (1969), *A Stochastic Response Model with Application to Brand Choice*, Management Science, Theory No. 7.

Mosteller, F. (1955), *Stochastic Models for Learning*, New York.

Nicosia, F.M. (1966), *Consumer Decision Process - Marketing and Advertising Implications*, Englewood Cliffs.

Opitz, O., and M. Schader (1975), *Operations Research Verfahren und Marketing Probleme* ,Mathematical Systems in Economics 19.

Pearl, R. (1925), *The Biology of Population Growth*, New York.

Pry, R.H. (1973), *Forecasting the Diffusion of Technology*, General Electric.

Rogers, E.M. and F. Shoemaker (1971), *Communications of Innovations*, New York.

Sahal, D. (1976), *The Multidimensional Diffusion of Technology*, Technological Forecasting and Social Change.

Stapleton, E. (1976), *The Normal Distribution as a Model of Technological Substitution*, Technological Forecasting and Social Change 8.

Topritzhofer, E. (1974), *Absatzwirtschaftliche Modelle des Kaufentscheidungsprozesses unter Berücksichtigung des Markenwahlaspektes*, Wien.

US Department of Commerce (1975), *Historical Statistics of the United States, Colonial Times to 1970, Part 1 and Part 2*, U.S. Department of Commerce, Bureau of Census, Washington DC.

Winsor, C.P. (1932), *A Comparison of Certain Symmetrical Growth Curves*, Journal of the Washington Academy of Sciences, Vol. 22.