Working papers of the N.I.D.I

Netherlands Interuniversity
Demographic Institute
P.O. Box 955
Prinses Beatrixlaan 428
2270 AZ Voorburg - the Netherlands
The Netherlands Interuniversity Demographic Institute - N.I.D.I. - was established in the Hague in 1970. Twelve scientific institutions of the Netherlands participate in it. In addition to the Royal Netherlands Academy of Sciences and Letters and the Interuniversity Foundation for Social Science Research (SISWO), these comprise the universities or specialized schools of Amsterdam (2), Delft, Leyden, Groningen, Nijmegen, Rotterdam, Tilburg, Utrecht and Wageningen. The universities and schools make equal contributions to the operating budget of the institute by means of subsidies granted by the Department of Education and Science. The N.I.D.I., directed by Prof. Dr. D.J. van de Kaa, is mainly concerned with research into the factors which, either directly or indirectly, influence demographic development in the Netherlands. The institute also tries to stimulate and co-ordinate such research in the member institutions, and maintains a close working relationship with the Central Bureau of Statistics and other data collecting and planning agencies.

The research interests of the N.I.D.I. are however, not restricted to the Netherlands. The Institute plans to undertake comparative research in European countries as well as research in developing countries. Where possible such research will be carried out in collaboration with institutes in the countries concerned.

Another important aim of the N.I.D.I. is to further the training of demographic experts. To achieve this aim the Institute supervises students during periods of practical work, provides guest lectures and will offer courses at the post-graduate level.

As an interuniversity institute the N.I.D.I. also has tasks with respect to the collection of research materials and documents. A library attuned to the needs of the staff and the research workers in the member institutions exists and will be developed further. A series of monographs in Dutch is published by contract with Van Loghum Slaterus (Deventer); a selected annotated bibliography in English of population studies in the Netherlands is prepared yearly in collaboration with the Netherlands Demographic Society. Another part of the Institute's publication programme results from close collaboration with the Flemish section of the Centre for Population and Family Studies in Brussels. The periodical "Benvolking en Gezin" (Population and Family) is published in Dutch; yearly three numbers are issued. An English language publication entitled "Population and Family in the Low Countries" is issued biennially by the Social Sciences Division of Martinus Nijhoff (The Hague) in the series NIDI/CBSSS-publications and contains translations of selected papers originally published in Dutch.

A Reprint Series and a Series of Working Papers, both begun in 1973, intend to facilitate the distribution of selected papers originating in the N.I.D.I. or prepared as part of the institute's programme of work.
This paper is part of the research project "Development of a methodology for regional population forecasting and for the study of population distribution policy". The project is carried out jointly by NIDI and the Delft University of Technology and is funded by the Physical Planning Agency (RPD).


** We would like to thank Nico Keilman and Joop de Beer for commenting on an earlier draft. We are also thankful to Joan Vrind, who expertfully typed the manuscript.

Multidimensional forecasting.
A systems approach *

by Frans Willekens and Nazli Baydar **

Working Paper no. 40

Voorburg, August 1983
ABSTRACT

Multidimensional forecasting is a general term for the forecasting of inter-related elements. Multivariate forecasting, vector forecasting and multi-item forecasting are members of this family. The interrelated elements constitute a system and forecasting the system is itself a process. The following steps of the forecasting process are discussed: systems description, systems measurement, systems analysis, systems modelling and systems forecasting. The description, measurement and analysis of the system proceed and simplify the selection of the forecasting model. Several models are reviewed and related to observation plans of the systems dynamics. The state-space model is the preferred forecasting model since it provides a basis for a unified perspective on multidimensional forecasting. Several conventional forecasting models may be formulated in the state-space format. Exponential smoothing and other techniques are presented to adjust the coefficients of the state-space model during the forecasting process.
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>ABSTRACT</td>
<td>III</td>
</tr>
<tr>
<td>1. INTRODUCTION</td>
<td>1</td>
</tr>
<tr>
<td>2. STEP 1: SYSTEMS DESCRIPTION</td>
<td>5</td>
</tr>
<tr>
<td>3. STEP 2: SYSTEMS MEASUREMENT</td>
<td>8</td>
</tr>
<tr>
<td>4. STEP 3: SYSTEMS ANALYSIS</td>
<td>14</td>
</tr>
<tr>
<td>4.1. Validation of systems description</td>
<td>14</td>
</tr>
<tr>
<td>4.2. Dynamics of system elements and of their linkages</td>
<td>15</td>
</tr>
<tr>
<td>5. STEP 4: SYSTEMS MODELLING</td>
<td>21</td>
</tr>
<tr>
<td>5.1. E-models: distribution techniques</td>
<td>22</td>
</tr>
<tr>
<td>5.2. ER-models: state-space models</td>
<td>27</td>
</tr>
<tr>
<td>5.2.1. Model identification</td>
<td>27</td>
</tr>
<tr>
<td>5.2.2. Parameter estimation</td>
<td>31</td>
</tr>
<tr>
<td>6. STEP 5: SYSTEMS FORECASTING WITH VARIABLE COEFFICIENT MATRICES</td>
<td>34</td>
</tr>
<tr>
<td>7. CONCLUSION</td>
<td>42</td>
</tr>
<tr>
<td>NOTES</td>
<td>45</td>
</tr>
<tr>
<td>REFERENCES</td>
<td>46</td>
</tr>
<tr>
<td>LIST OF WORKING PAPERS</td>
<td></td>
</tr>
</tbody>
</table>
1. INTRODUCTION

Multidimensional forecasting is a general term for the forecasting of interrelated elements. Multi-item forecasting and multivariate forecasting are members of this family. In multi-item forecasting, the focus is on the simultaneous projection of each item belonging to a group, i.e. a meaningful and predictable aggregate (see e.g. Steece and Wood, 1979; Thomopoulos, 1980, p. 274). The items may be thought of as categories of a particular group-variable. In multivariate forecasting, the focus is on the simultaneous projection of interrelated variables (see e.g. Jenkins and Alavi, 1981). These variables may or may not be decomposed into categories. The label multidimensional forecasting was used by Thomopoulos (1980, p. 267). Sometimes reference is made to vector forecasting to express the relatedness of the items to be forecasted. This term indicates an analogy to the term vector time series, which is sometimes used to denote multiple time series (e.g. Granger and Newbold, 1977, p. 215) and to the vector smoothing, used by Brown (1962, pp. 199-206) to denote the application of exponential smoothing to a probability distribution of demands. Vector forecasting is used when the interrelated items are arranged in a vector. The items may, however, also be arranged in a matrix or in a n-dimensional array. Multidimensional forecasting is a general term to denote the simultaneous forecasting of interrelated elements arranged in a n-dimensional array.

The categories (items) and/or variables are assumed to be related to one another, such that a change in one component of the whole is affected by and affects variations in other components. The interdependence warrants and even requires simultaneous consideration of all components. Multi-item forecasting arises in several disciplines. Demographic forecasts are generally multi-item forecasts: the items are represented by age groups or other population characteristics such as marital status, labor force status (in labor supply projections), region of residence (in regional population projections) or household category (in household projections). It is therefore not surprising that demographers are paying much attention to methods for simultaneously forecasting different population categories. Research on the analysis of interdependent subpopulations gave rise to the new field of multidimensional demography (Rogers, 1975; Land and Rogers, 1982). The simplest illustration of multi-item forecasting in economics may be found in the input-output literature, which focusses on the interdependence among sectors of the economy. Input-output models in general and dynamic input-output models in particular are designed to analyse and forecast the structure of a multi-sector economy (see e.g. Almon et al., 1974; Blin et al., 1979). Other illustrations are the forecasting of the consumption of related goods (e.g. dairy products). An illustration in management science is provided by Steece and
Wood (1979). They develop an ARIMA-based methodology for predicting individual item demand for classified inventory and apply the method to forecast the demand of three types of antihistamines, a typical hospital pharmacy inventory problem.

Multivariate forecasting arises predominantly in economic forecasting. Illustrations are the simultaneous projection of employment and migration, rate of inflation and rate of unemployment, price and quantity, etc. In this paper, we are particularly interested in multivariate situations in which the relations between any pair of variables runs in two directions (direct relation and feedback).

The need for multidimensional forecasts in several disciplines did not yet result in a unified perspective. Recent developments in multivariate time series methods do not contain this perspective, as will be shown in this paper. Better suited are the state-space models, which offer a flexible basis for a unified perspective on multidimensional forecasting. This paper aims to outline a systems approach to multidimensional forecasting. A systems approach, we believe, will provide a unified multidimensional forecasting framework. Two observations are at the origin of this paper. First, recent developments in demographic forecasting models may be of use to predict individual items in a multi-item environment. This observation is an outcome of our reading of the literature on multivariate time series, as part of our research towards the development of an improved demographic model for forecasting interrelated population categories. The second observation relates to the problems encountered in representing and forecasting structural change. In this paper structural change is defined as a change in the interrelatedness among the items or variables. Since, in general, the items considered in multi-item forecasting are categories of a categorical or nominal variable, recent developments in categorical data analysis may be of use to represent and forecast structural change. Of particular interest may be the odds-ratio concept, which is a measure of association or interrelatedness, and the log-linear model, which may be used to describe a complex multivariate time series of categorical data in terms of interaction effects. Both observations led to the ideas developed in this paper and to the unified forecasting framework.

The paper describes the forecasting process. Because of the interrelatedness of the individual items and variables, the aggregate that contains the items and variables is treated as a system. The items constitute the elements of the system. The variables constitute subsystems. This systems approach to multidimensional forecasting enables us to easily clarify several issues and to suggest ways to draw on the mathematical systems theory for the development of a forecasting model. The advantage of the systems approach for the analysis, projection and control of classified items was demonstrated by Willekens (1976) in a study of multiregional population systems.
About six steps may be distinguished in forecasting a system:

1. systems description
2. systems measurement
3. systems analysis
4. systems modelling
5. systems forecasting (hypothesis formulation)
6. evaluation and monitoring of forecasts (error analysis; acceptance by user; etc.).

The steps describe the forecasting process. In this perspective on forecasting, the development of the projection model is no longer the only dominant aspect of forecasting. Modelling is embedded in a broader framework. The shift in attention from forecasting techniques to a forecasting process is also apparent in some recent publications (e.g. Makridakis and Wheelwright, 1982). The forecasting process itself is embedded in a broader framework. Jenkins (1982, p. 4) recently emphasized that forecasting is a means to aid decision making and is not an end in itself. He stated: "Forecasting systems frequently go wrong not only because of the poor technical quality of the forecasts but also because insufficient attention has been paid to the relationship between forecasting and decision taking" (Jenkins, 1982, p. 4).

Makridakis and Wheelwright observed that "One of the most interesting developments in the field of forecasting in the late 1970's was the realization that forecasts alone are useless until applied for planning and decision making purposes" (Makridakis and Wheelwright, 1982, p. 556). This view is echoed by Fildes who states that "The key to evaluating the organization's forecasting performance is to examine how forecasts are used - not just how they are produced" (Fildes, 1982, p. 86). Most pitfalls in forecasting are situated on the interface between forecaster and the organizational environment and are frequently caused by inadequate communications between the forecaster and the user (planner, manager, decision-maker). Hogarth and Makridakis (1981) and Armstrong (1982) review several of the pitfalls that arise by an inadequate integration of forecasting and decision making. This observation induced authors to propose integrated forecasting-decision making systems. One such system was presented by Jenkins (1982, p. 5). A similar system was shown by Fildes (1982, p. 85) (Figure 1). Therefore, the forecasting process should be thought of as being part of a decision support system. Forecasting involves anticipating future events; planning (decision making) is concerned with devising courses of action to influence the occurrence of future events. Forecasting is an essential component of strategic planning. In this context questions such as "who needs the forecasts ?", "why is the forecast for ?", "who will be using the result and in which way ?" and "how does the system for which
the forecast is designed operate and which are the institutional constraints?" become very important. They may be dealt with by a proper organization of the forecasting process, which includes the design of adequate information channels. It should have become clear by now that forecasting is much more than the development and application of a projection model. Forecasting is not a technique; it is a technology.

Steps 1 to 4 of the forecasting process will be covered in this paper. Some aspects of step 5 will also be discussed, namely the forecasting with variable coefficient matrices. The major problem of hypothesis formulation will, however, not be dealt with in this paper. The paper attempts to show that some major attempts to develop multidimensional forecasting models as generalizations of univariate models failed because of system measurement problems. Traditional observation plans are ill-suited for multidimensional forecasting.

A forecasting model that allows to simultaneously forecast not only the individual items of one group (categories of one variable) but also the items belonging to different groups (categories of several variables, constituting a multi-way classification), is referred to as a multidimensional forecasting model. Of key concern in multidimensional forecasting is the interrelatedness or interdependence of all classified items and of the variables in the classification. A basic issue in modelling is therefore the appropriate description of the observed interdependencies and the prediction of changes in interdependencies. These changes represent a structural change. The forecasting approach that will be presented, emphasizes the interdependencies and their patterns of change. The classified items are the outcome of underlying interactions between the variables in the classification.

The general discussion up till now regarding forecasting led to two propositions, that are essential for the further discussion in this paper. The two propositions are:

(i) forecasting is a process

(ii) multi-item forecasting and hence multidimensional forecasting may be enhanced by adopting a systems analytic perspective.

In the introduction, the forecasting process was considered to be composed of six steps. We will now discuss steps 1 to 4 and some aspects of step 5.
2. **STEP 1: SYSTEMS DESCRIPTION**

Because of the interrelatedness of the items in a multi-item forecasting problem, the classified items may be treated as elements of a system (the aggregate). A system is simply a set of interrelated elements. The systems concept emphasizes the interrelatedness or interdependence of the elements that belong to a particular system. Elements not included in the system, although they have some relation to elements in the system, are said to belong to the environment of the system. Within a system, two components are always present: the elements (items) and the relations between the elements. Frequently, subsystems may be distinguished, namely, if clusters of elements can be identified with considerably more within-cluster interaction than between-cluster interaction. Recall that different variables (with or without categories) are also considered to constitute different subsystems. The system may therefore be composed of a hierarchy of subsystems.

The first step of the forecasting process is the identification of the system to be forecasted. This step consists of two sub-steps. First the description of the static properties of the system at a reference point in time \( t \) or a reference period \( (t, t+1) \). Second, the description of the dynamic properties, i.e. the relevant patterns of change and the identification of the underlying components of change.

a. The static system.

The elements are classified items and are denoted by \( i \) \( (i = 1, 2, ..., N) \), where \( N \) is the total number of elements in the system). They constitute the units of analysis. The classification should be meaningful to the user of the forecast and should be suited for forecasting. These two objectives are sometimes conflicting. For instance, a meaningful age grouping of the population is derived from the stages in the life cycle. Unequal age intervals are, however, problematic in forecasting, since no demographic projection model adequately processes unequal age intervals. Sufficient attention to a forecasting-oriented classification scheme may considerably simplify the other steps of the forecasting process. Two criteria are of predominant importance for the design of forecasting-oriented classification schemes:

(i) homogeneity: the system elements (categories) should be as homogeneous as possible in its composition and

(ii) stability: the individual elements should have a regular pattern of change and, therefore, be easy to predict.

**Relations** between elements are denoted by a combination of elements. For instance, \( (i, j) \) denotes the relation between element \( i \) and element \( j \). Relations are
not always easy to identify and are frequently difficult to measure and hence to quantity. Ideally, relations result in transfers or transactions between the elements. The transactions may be identified in terms of flows of people, goods, capital or information. Transactions are the realizations of relations between elements. The relations essentially determine the structure of the system.

b. The dynamic system.

Systems description for forecasting also involves the identification of the patterns of change of elements and relations. In most time-series models, within autoregressive moving average scheme, relations between system elements are not identified. In this case, the dynamic system is simply described by the elements at successive points in time. Sometimes, the relations are inferred from a time series of descriptions of elements. This indirect identification (and measurement) of relations is characteristic for many econometric and time series models. This point will be clarified later, since it is the basis for the derivation of the necessary features of models that are suited for multidimensional forecasting. Changes in elements are due to increments and decrements brought about by the relations and changes in the relations. The transactions between system elements and between elements and the environment constitute the basic events that cause the elements to change. Ideally, the description of the systems dynamics should therefore focus on events, such as migrations from one region to another; switches from one brand of cigarette to another; transactions between sectors of the economy; relocations, births, deaths or merges of companies; reallocation of funds, etc. In some forecasting models, the relations and the events that generate change have become the focus of attention. Examples are demographic models, input-output models and systems dynamics (Forrester) models.

The emphasis on relations (and on their realizations: transactions or events) is a key feature of this paper. We will therefore also pay attention to changes in relations. We define structural change as change in the relations between the elements. Changes in the elements themselves and/or in the significance of each of the elements in the whole system are not considered as structural changes. For instance, changes in population distribution patterns or market shares are not qualified as structural changes if they are the outcome of, respectively, migration patterns and shifts of consumers between markets.

The outcome of the systems description step may be organized in a way to enhance the implementation of the other steps. Recall that the system may consist of subsystems defined by variables, elements defined by categories of the variables, and relations among the categories and/or variables defined by
transactions. A dynamic system may ideally be represented by an accounting framework. Many relations are not directly observable and can therefore not be represented in an accounting framework. In such cases, one is left with a statistical framework, in which relations are inferred or estimated. Since the framework is not independent of the degree to which the system components can be measured, it will be presented in the next section.
3. **STEP 2: SYSTEMS MEASUREMENT**

A system consists of elements and relations. The system components may have certain characteristics associated with them, which are represented by a variable. For instance, in a multiregional system, the individual elements (regions) may be characterized by their population size, number of births, level of export, total gross regional product, number of physicians, propensity to invest, export quote, birth rate, etc. The characteristics may be expressed in absolute terms or in relative terms. Each characteristic of the elements imply a particular characteristic of the relations.

Systems measurement is the measurement of characteristics of elements and relations and hence the generation of values for the variables. It is the data collection phase of the process. Measurement involves an observation plan or measurement procedure. The observation plan has an effect on all steps in forecasting, but in particular on systems modelling: the measurement procedure and the model selection are tightly connected. Although the literature contains statements such as "the type of method depends on the type of data", the real significance of measurement for the selection of forecasting methods is often neglected. Within the framework of time series methods, often the quantity of available data is emphasized as relevant to method selection and not the type of available (or desirable) data. The observation plan describes which data are collected, how they are collected (e.g. sampling), the quality of the data and how the data relate to the forecasting model.

The ideal observation plan for dynamic systems modelling and hence for forecasting consists of complete, continuous-time observations on all elements and relations. In practice, this is generally impossible. It may even not be feasible to measure the relations at all and the forecaster must rely on a time series of element measurements only. In the context of forecasting, two broad groups of observation plans may be distinguished. With each group is associated a distinctive modelling perspective. The first group consists of observation plans, that are limited to the elements (**E-observation**). In general, the data relate to characteristics of the element, measured in regular intervals (discrete time). Let the state of element i at time t be denoted by $x_i(t)$. The outcomes of the observation at time t may be combined in a vector $x(t)$ with elements $x_i(t)$. The vector $x(t)$ is the state vector; it describes the state of the system at time t. In statistics the E-observation plan is also referred to as panel observation. E-observations are suited for forecasting on the basis of time-series methods.

The second group consists of observation plans in which information on the elements as well as on the relations are collected (**ER-observation**). Recall that the
relations are generally expressed in terms of transactions or transfers between the elements. Let $a_{ij}(t)$ be the transfer between element $i$ and element $j$ at time $t$ or in the interval from $t$ to $t+1$.

The outcomes of the observations on elements and relations may be arranged in an accounting framework. The account presents the state of the system at two consecutive points in time and the transactions in the interval between the two points. Table 1 presents such an account for the time period $(t, t+1)$. The system is assumed to consists of three elements. The environment of the system is denoted by $e$. Note that in a closed system the total size remains constant $\sum_i x_i(t) = \sum_i x_i(t+1)$ and the only dynamics is a redistribution. Illustrations of complete accounting frameworks are demographic accounts (e.g. Rees and Wilson, 1977), social accounts (United Nations, 1975; Juster and Land, 1981) and economic (input-output) accounts (e.g. Leontief, 1941) and economic-demographic accounts (Batey and Madden, 1981). All these accounts are similar to Table 2. By way of illustration, the Batey-Madden framework is shown in Table 3. If the relations between system elements cannot be observed, the account may be simplified substantially. The account is given in Table 4. Each column may be denoted by a vector. The table shows then a vector time series. The table contains the basic information for conventional time-series analysis and methods.

The measurement of all transfers (events) requires continuous time observation. Unless a registration system exists, this is impossible and a simplified observation plan must be designed. Following one observation plan, a transfer is recorded if it took place within some time prior to the enumeration (survey) date. This observation plan is illustrated in input-output surveys and may also be found in labor-force surveys and household surveys. In another observation plan, transfers are measured indirectly by comparing the state of the system at one point in time with the state at a previous point in time. This observation plan will be referred to as a discrete-time observation, since the comparison is between the beginning and the end of a time interval. The latter observation plan is a common source of migration data in Anglo-Saxon countries and is common in labor force surveys. The information on transfers is derived by comparing the region of residence or labor-force status at the time of enumeration (census or survey) with the region of residence or labor-force status one or five years prior to the census or survey. The measurement of relations among system elements is generally very costly and the high costs prevent a continuous-time observation and sometimes even an observation at regular intervals (discrete time observation). In a few fields, such as in input-output analysis, non-survey techniques are developed to generate up-to-date information on transfers on the basis of a full transfer matrix of a base period and recent aggregate information (Round, 1978; Conway 1983). Non-survey
Figure 1. The forecasting system

-Diagram description-

Direct influences or flows. The shaded boxes represent the major components of the forecasting system. The bold-outlined boxes represent actions taken by the participants—the forecaster and the decision maker. The broken boxes represent information and assumptions.

Table 1. Multi-regional demographic account

<table>
<thead>
<tr>
<th>ORIGIN</th>
<th>Regions of country</th>
<th>Abroad</th>
<th>Death</th>
<th>Total (t)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1</td>
<td>K_{11}</td>
<td>K_{12}</td>
<td>K_{1N}</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>K_{21}</td>
<td>K_{22}</td>
<td>K_{2N}</td>
</tr>
<tr>
<td></td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td></td>
<td>N</td>
<td>K_{N1}</td>
<td>K_{N2}</td>
<td>K_{NN}</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>inflow</th>
<th>inflow</th>
<th>outflow</th>
<th>outflow</th>
</tr>
</thead>
<tbody>
<tr>
<td>Abroad</td>
<td>K_{01}</td>
<td>K_{02}</td>
<td>...</td>
<td>K_{0N}</td>
</tr>
<tr>
<td>Birth</td>
<td>K_{B1}</td>
<td>K_{B2}</td>
<td>...</td>
<td>K_{BN}</td>
</tr>
</tbody>
</table>

| Total (t+1) | K_{+1} | K_{+2} | ... | K_{+N} | K_{+0} | K_{+δ} | K_{++} |
| inflow      |        |        |     |        |        |        |        |
Table 2. Complete (ER) accounting framework for multidimensional forecasting (time reference \( t \); three system elements)

<table>
<thead>
<tr>
<th>Origin of transactions</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>e</th>
<th>State of system at ( t+1 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Destination of transac-</td>
<td>( a_{11}(t) )</td>
<td>( a_{21}(t) )</td>
<td>( a_{31}(t) )</td>
<td>( a_{e1}(t) )</td>
<td>( x_{1}(t+1) )</td>
</tr>
<tr>
<td>tions</td>
<td></td>
<td></td>
<td></td>
<td>e</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>( a_{12}(t) )</td>
<td>( a_{22}(t) )</td>
<td>( a_{32}(t) )</td>
<td>( a_{e2}(t) )</td>
<td>( x_{2}(t+1) )</td>
</tr>
<tr>
<td>3</td>
<td>( a_{13}(t) )</td>
<td>( a_{23}(t) )</td>
<td>( a_{33}(t) )</td>
<td>( a_{e3}(t) )</td>
<td>( x_{3}(t+1) )</td>
</tr>
<tr>
<td>e</td>
<td>( a_{1e}(t) )</td>
<td>( a_{2e}(t) )</td>
<td>( a_{3e}(t) )</td>
<td>--</td>
<td></td>
</tr>
<tr>
<td>State of system at ( t)</td>
<td>( x_{1}(t) )</td>
<td>( x_{2}(t) )</td>
<td>( x_{3}(t) )</td>
<td>--</td>
<td>( x_{t}(t+1) )</td>
</tr>
</tbody>
</table>

Table 3. The basic structure of the demographic-economic forecasting framework

Table 4. Simplified (E) accounting framework for multidimensional forecasting: state of system at consecutive points in time

<table>
<thead>
<tr>
<th>Element</th>
<th>t</th>
<th>t+1</th>
<th>t+2</th>
<th>......</th>
<th>t+T</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>x₁(t)</td>
<td>x₁(t+1)</td>
<td>x₁(t+2)</td>
<td>......</td>
<td>x₁(t+T)</td>
</tr>
<tr>
<td>2</td>
<td>x₂(t)</td>
<td>x₂(t+1)</td>
<td>x₂(t+2)</td>
<td>......</td>
<td>x₂(t+T)</td>
</tr>
<tr>
<td>3</td>
<td>x₃(t)</td>
<td>x₃(t+1)</td>
<td>x₃(t+2)</td>
<td>......</td>
<td>x₃(t+T)</td>
</tr>
<tr>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td></td>
<td>.</td>
</tr>
<tr>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td></td>
<td>.</td>
</tr>
<tr>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td></td>
<td>.</td>
</tr>
<tr>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td></td>
<td>.</td>
</tr>
<tr>
<td>N</td>
<td>xₙ(t)</td>
<td>xₙ(t+1)</td>
<td>xₙ(t+2)</td>
<td>......</td>
<td>xₙ(t+T)</td>
</tr>
</tbody>
</table>

| Total   | x⁺(t)  | x⁺(t+1)| x⁺(t+2)| ...... | x⁺(t+T) |

Techniques have a great potential in updating historical data on transfers among system elements (for a methodological review, see Willekens, 1982a).

The differences between continuous-time data and discrete-time data on transfers call for differential modelling of the systems dynamics. The link between observation plan and systems modelling has extensively been studied for multiregional demographic projection models, where the transfers are represented by migration and multidimensional labor force projection models with transfers being changes of labor force status (e.g. employed, unemployed, out-of-labor-force) (Willekens, 1982b). ER-observations allow a straightforward application of state-space models for forecasting since the transition (coefficient) matrices are directly observable.

The adoption of an accounting framework has a number of advantages. First, the account gives an overview of the available data. The flow data (data on transfers) and the stock data (data on state of system) are neatly arranged. Second, if all data are entered in the account, the account must balance. This condition allows easy identification of inconsistencies in the data set. Third, the account provides a useful framework for estimating missing data and for updating historical data. Finally, the variables and parameters of the state-space forecasting model may easily be derived from the account. The state variables are the $x_i(t)$'s and the transition probabilities are calculated from the $a_{ij}(t)$ values. Note that in systems dynamics (Forrester) models stock variables are denoted by level variables and flow variables by rate variables.
4. **STEP 3: SYSTEMS ANALYSIS**

Forecasting without resort to explanatory (econometric) models is based on the premise that the future is (partly) contained in the past. The analysis of past and present dynamics of the system is therefore prerequisite to good forecasting. The systems analysis comprises two aspects: first, the validation of the systems description and second the investigation of the pattern of change of the system elements and of the relations or linkages between the elements. The analytical techniques used to perform the analysis may range from simple methods for exploratory data analysis to sophisticated models. In this section, we forego on the technical tools for analysis but focus instead on the problem areas of forecasting-related systems analysis.

4.1. **Validation of systems description**

In the discussion of the systems description, we claimed that the identification of elements on the basis of forecasting-oriented classification schemes may significantly simplify the other steps of the forecasting process. To be forecasting-oriented the elements should satisfy the conditions of homogeneity and stability. The homogeneity condition requests the individual items in a category, i.e. system elements, to be as homogeneous as possible. The condition is derived from the assumption, implicit in any systems model, that all items belonging to a particular category have the same behavior, represented by a unique set of parameter values. For instance, in age-specific demographic growth models, the system elements consist of age groups and all individuals in a particular age group are assumed to have the same mortality and fertility rate, which is the weighted average of the rates pertaining to the homogeneous sub-groups; in input-output models, all corporations in a particular sector are assumed to have the same input coefficients. If an element (category) is not homogeneous, the same set of parameter values may not be applied to all items contained in the element. There are two ways to deal with heterogeneity. The first and simplest way is to increase the number of elements in the system (and hence the state space) until relatively homogeneous elements are obtained. A classic illustration of this approach is the mover-stayer model of migration (see e.g. Bartholomew, 1973). The second approach does not increase the number of elements and hence the number of parameters. Instead, a parameter pertaining to a particular element is expressed as a function of the characteristics of the individual items. Heckman and Singer (1982) discuss this approach in the context of demographic models.
The second condition to be satisfied by forecasting-oriented classification schemes is that the elements are stable, i.e. have a regular pattern of change. If all elements are stable, forecasting is simply a matter of carrying the constancies into the future. An important aspect of the systems analysis is therefore the search for stability. Two questions should be addressed in the search for stability:
- which characteristics of the system elements are stable?
- what is the degree of stability?

The problem of identifying a stable pattern in a sequence of observations $x_i(t)$ relative to the characteristics of a particular element is very much analogous to the well known time series analysis techniques which aim at detecting the stochastic noise from the underlying pattern of the time series. Two main branches of time series analysis techniques may be distinguished: autocorrelation analysis and spectral analysis (Makridakis, 1976, 1978). Autocorrelation analysis aims at identification of stationarity, seasonality and types of generating processes, where spectral analysis aims at identification of frequency response characteristics of the generating process.

4.2. Dynamics of system elements and of their linkages

In the section on systems measurement, we distinguished between the two broad groups of observation plans: E-observation and ER-observation. With each group is associated a distinctive modelling perspective. The E-observations yield the data base for multivariate time-series models; the ER-observations provide the data for a direct estimation of the coefficient matrices of state-space models. It will be shown that the selection of the appropriate model within each perspective should depend on the outcome of the analysis of the relatedness of the system elements, of the linkages and of their patterns of change. We discuss separately the data from E-observations (time series of $x(t)$-vectors) and the data from ER-observations (time series of $A(t)$-matrices).

a. System elements.

In section 4.1 we have discussed the search for stable components (analogous to time series analysis techniques) in the pattern of change of the system elements. The time path of an element or item in the classification was considered independently of the time path of the other elements. In this section we focus on the patterns of change of all elements simultaneously and search for stability in the relative position of an
element in the whole (or aggregate); in other words, for the stability of the frequency distribution of the aggregate. To clarify the distinction between the search for stability discussed in the previous section and the one to be discussed in this section, note that the elements $x_i(t)$ may be arranged in a matrix with rows $i = 1, 2, ..., N$ and columns $t = 1, 2, ..., T$ (Table 2). The investigation of the time path of an element focuses on each row of the matrix separately. The study of the relatedness of elements focuses on the columns and on the changes of the columns over time.

Recall that the characteristics of the elements may be measured in absolute terms or in relative terms. It is assumed here that the characteristics are measured in absolute terms (e.g., population size; total demand) and that the relative position of an element is expressed in terms of share or fraction of the aggregate. For a given characteristic, the time series of the absolute values for the elements are completely determined by the aggregate series and the series of fractions. In general, the aggregate series and the series pertaining to individual elements or items are more or less correlated. The measurement of the degree of correlation (association) requires measures of association. A well-known example of these types of measures of association within the time-series analysis framework is the cross autocorrelation function, which is the multivariate generalization of autocorrelation. At this point an important remark may be made. A correlation between the aggregate series and the component series means that both the aggregate and the elements have similar time dependencies. The time dependency of each of the elements or of the aggregate may be studied by describing the time series by a log-linear model. This approach will be elaborated in step 4 and is the basis for the unified framework for multidimensional forecasting.

In the case of a perfect correlation between the series of absolute values for each individual element and for the aggregate, the fractions are stable. It suffices to forecast the aggregate and to apply the constant fractions at each time to obtain forecasts for the individual elements. Note that in this case the correlation between the aggregate series and the fractional series (series of fractions) is zero. The approach of forecasting the aggregate and of deriving the elements by some fraction or distribution function is referred to as the top-down approach. The class of models following the top-down approach will be discussed later.
If for a particular characteristic of the system elements the time series of the absolute values for each individual element and for the aggregate are not correlated, but are independent then the top-down approach to forecasting is inappropriate. The patterns of change of the elements may differ because of differences in the elements and/or because they react differently to exogenous influences. In the case of independence of the individual element series, a better approach to forecasting is to forecast each element separately. The forecast of the aggregate is then the sum of the element forecasts. This approach is referred to as the bottom-up approach.

b. Relations or linkages between system elements.

In the previous section, the focus was on the stability of the relative position of each of the elements in the whole (aggregate), i.e. on the stability of the elements $x_i(t)$ of the state vector $x(t)$ if compared to the aggregate $\sum x_i(t)$. This section focuses on the search for stability of the relation between the system elements. The data for the stability analysis are the transfers between the elements, i.e. $a_{ij}(t)$ for all $i, j$ and $t$. They may be arranged in a three-dimensional table with rows $i = 1, 2, \ldots, N$; columns $j = 1, 2, \ldots, N$ and layers $t = 1, 2, \ldots, T$ (Table 2). Measurement problems, including the cost of measurement, frequently make the availability of a time series of transfer data impossible. However, data limitations should not preclude the desirability of data on relations, or attempts to develop alternative, feasible, indirect measurement and estimation techniques. Due to measurement problems the pattern of change of the relations usually cannot be studied and consequently the state-space model becomes time invariant (i.e. has constant coefficients).

The existence and the relative size of the linkages between system elements define the structure of the system. Structural analysis of the system consists of a study of the linkages and of their patterns of change (i.e. their time-dependence, or time-heterogeneity). We will refer to structural forecasting to denote the forecasting of the relative position of linkages in the system.

To study the correlation of linkages between system elements, it may be useful to distinguish between two sets of linkages. The first set comprises different types of linkages (and the flows that result) between the same two elements. The second set comprises linkages of a particular type between different pairs of elements. An illustration of the first set is the
flow of capital, jobs and people between two regions. If these flows are
dependent, they should be treated simultaneously.

The structural analysis of relations among linkages calls for the
development of measures of association. Two measures of association are
presented: the odds ratio and the elasticity. Other measures of association,
related to the odds ratio, are discussed by Altham (1970) and Clogg (1979,
Appendix E).

(i) odds ratio: the odds is a ratio of probabilities or frequencies and the
odds ratio (cross-product ratio) is the ratio of odds; odds ratios may be
calculated for any 2 x 2 subtable of the N x N x T-table. The odds
ratios centered around the element \( a_{ij}(t) \) are:

\[
\Omega^{(1)}_{ijt} = \frac{a_{ij}(t)}{a_{i+1, j}(t)} / \frac{a_{i+1, j+1}(t)}{a_{i+1, j+1}(t)}
\]

\[
\Omega^{(2)}_{ijt} = \frac{a_{ij}(t)}{a_{i+1, j}(t)} / \frac{a_{ij}(t+1)}{a_{i+1, j}(t+1)}
\]

If for a particular \( t \), the origins of the relations (transfers) and the
destinations are independent, \( \Omega^{(1)}_{ijt} = 1 \). The first type of odds ratios
therefore measures the association between origins and
destinations. The magnitude of the association is given by the deviation
of the odds ratio from one. Note that the average magnitude of
association between origin and destination is given by

\[
\Omega^{(1)}_{ij+} = \frac{a_{ij}(+)\cdot a_{i+1, j+1}(+)}{a_{i+1, j}(+)\cdot a_{i+1, j+1}(+)}
\]
where + denotes the sum of \( a_{ij}(t) \) over the subscript replaced by the +. The second type of odds ratio measures the time dependence of the transfers originating in \( i \) for a particular destination. The average time dependence of the transfers originating in \( i \) is

\[
\Omega_{i+t}^{(2)} = \frac{a_{i+}(t)}{a_{i+1+}(t)} / \frac{a_{i+}(t+1)}{a_{i+1+}(t+1)} = \frac{R_{i+1+}(t)}{R_{i+}(t)},
\]

which is equal to the ratio of the growth ratio of \( i + 1 \) to the growth ratio of \( i \).

Note that the logarithm of an odds ratio is the second-order difference of the logarithms of adjacent linkages.

(ii) elasticity. An elasticity is the percentage change in one variable or item relative to a percentage change in another (other things being constant). For instance, the elasticity of the linkage \( a_{ij} \) with respect to the linkage \( a_{kj} \) is

\[
\frac{a_{ij}(t+1) - a_{ij}(t)}{a_{ij}(t)} / \frac{a_{kj}(t+1) - a_{kj}(t)}{a_{kj}(t)} = \frac{r_{ij}(t)}{r_{kj}(t)} = k_{ij}^{\varepsilon}.
\]

The elasticity is a ratio of two growth rates. It may be written as follows:

\[
\left[ \frac{a_{ij}(t+1)}{a_{ij}(t)} \right] - 1 = \frac{R_{ij}(t) - 1}{R_{ij}(t)},
\]

which is the ratio of (odds - 1). Notice the difference with \( \Omega_{ij}^{(2)} \), which may be written as follows:
$$\frac{1}{\Omega_{ijt}^{(2)}} = \frac{a_{ij}(t+1)}{a_{ij}(t)} / \frac{a_{kj}(t+1)}{a_{kj}(t)} = \frac{R_{ij}(t)}{R_{kj}(t)}$$

with $k = i + 1$ and $j$ fixed.

The measures of association presented above quantify the relatedness among system elements. For each set of four linkages, an odds ratio may be calculated. Hence, the linkages between two pairs of elements may be summarized by the odds ratio. Another way to measure the relatedness of linkages is to describe the $a_{ij}(t)$-data in the three-dimensional table by a generalized linear model. The model represents an $a_{ij}(t)$-value in terms of a set of parameters, each of which quantifies a particular relatedness, exhibited by the ijt-relation with regard to the other linkages. The application of this type of models to the study of stability of linkages is illustrated by Baydar (1983) in her study of the stability of interprovincial migration patterns in The Netherlands.
5. **STEP 4: SYSTEMS MODELLING**

Once steps 1 to 3 have been studied in sufficient detail, the development of the forecasting model is relatively straightforward. The forecasting model is a model of the systems dynamics. As we have already mentioned, different classes of models are suited for forecasting. The "appropriate" model depends on the type of system to be forecasted, on the available data and on the outcomes of the preliminary analysis of the data. Model building is itself a process. Following the approach to model building proposed by Box and Jenkins (1976, p. 171), three stages can be distinguished:

a. model identification: identification of a class of models worthy further investigation and selection of a particular model, to represent the data,

b. model estimation: estimation of the parameters,

c. diagnostic checking (model validation): checking the fitted model to reveal model inadequacies and to achieve model improvement.

In this paper we focus on the model identification. In section 5.2.2 we discuss briefly parameter estimation for the ER-models. Model validation, however, will not be discussed. The distinction between model fit and its forecasting performance is the main issue at model validation phase. For a review see Makridakis (1976) and (1978). Only multivariate (or multi-item) models will be considered, since the subject of the paper is the simultaneous forecasting of variables, some of which may be categorical. It is shown that the identification of an appropriate model becomes relatively easy if steps 1 to 3 have been carefully considered.

Two major classes of multivariate forecasting models were distinguished in section 2.2 of the paper. The choice between the classes are largely determined by the observation plan (i.e. the data). E-observations lead to conventional multiple time-series models. Multiple time series models forecast the state of a system, expressed by the state vector $x(t)$, on the basis of a time series of state vectors. This class of models will be referred to as **E-forecasting models**. The basic feature of E-models is that they forecast the system (set of elements, i.e. variables and/or items) on the basis of the knowledge of the state of the system at consecutive points in time only. ER-observation plans enable the construction of more sophisticated multivariate forecasting models. These models express changes in the state vector in terms of the underlying linkages between system elements. They are referred to as **ER-models** and take the format of state-space models. In an ER-observation plan, the linkages between the system elements are directly observed. Frequently, however, direct measurement of the linkages is not feasible, although their existence is beyond doubt.
In this case, the linkages may be inferred from a time series of state vectors and/or from the incomplete information one might have on the linkages. The format of these models may not necessarily be the state-space format. They can however be transformed into a state-space format.

5.1. **E-models: distribution techniques**

E-models describe the system dynamics in terms of changes in the system elements only. In section 2.3 two classes of models were distinguished. In top-down models the multiple-time series of characteristics of elements are written as a time series for the aggregate and a set of time-series of fractions. The fractions represent the share of an individual item in the aggregate. The main advantage of top-down models is that they assure consistency between the individual item forecasts and the aggregate forecast. The second class, the bottom-up models, represents a multiple time series by a set of independent time series, one for each item. The aggregate time series is the sum of the individual time series. Bottom-up models enable to fully account for the particular features of each item. The main disadvantage, however, is the inconsistency between the sum of the individual series and a given aggregate series. A third class of mixed models may be added.

The selection of the appropriate type of forecasting model, that is associated with E-observations, should be based on the analysis of the relatedness of the system elements, i.e. the analysis of the $x_i(t)$-series for all $i$ (see Step 3). If the series exhibit a high correlation, the top-down approach to multidimensional forecasting is appropriate: the multivariate or multi-item series is represented by a univariate series of the aggregate and a vector series of fractions. If the series are independent, the bottom-up approach is preferred: the multivariate or multi-item series is represented by a set of independent univariate series. The top-down and bottom-up approaches are of course the extremes of a range of approaches which combine features of the top-down and the bottom-up approaches. The simplest illustration of mixed approaches is the projection of the individual elements separately and then to adjust the results to assure consistency with the aggregate time series. An alternative, equally simple approach, is to adopt a top-down approach but to adjust the fractions to account for changes in the fractions. Such a procedure was applied by Steece and Wood (1979). The authors use an ARIMA model to forecast the aggregate and adjust the fractions by a single exponential smoothing. Illustrations of more complex mixed approaches are found in the literature on demographic forecasting.
Keyfitz (1982) discusses an illustration in the context of mortality. The elements are age groups and the characteristic to be forecasted is the survival probability pertaining to the age group. Several alternatives may be imagined. Each survival probability may be forecasted separately. Only their sum, which may be expressed as expectation of life at birth may be forecasted. Alternatively the overall pattern of the probabilities expressed by a few parameters may be forecasted. These alternatives constitute examples of bottom-up, top-down and mixed approaches respectively. From the forecasted pattern parameters the age specific probabilities of survival are generated. The issue of modelling multiple time series using a reduced parametrization was recently also discussed by Reinsel (1983).

Top-down models distribute, at each time, the forecasted aggregate among the elements in the multiple time series. In doing this, some measure of interrelatedness of the elements is held constant. Each top-down model therefore implies a particular constancy. This observation is the clue to link systems modelling to systems analysis. The optimal forecasting model is determined by the outcome of the search for stability. The forecasting model carries some constancies into the future. The challenge is to discover in past data, the constancies that could be extended to predict the future (for a philosophical note, see Makridakis, 1976, p. 62). In this sub-section we consider only deterministic models.

Top-down models have extensively been used in demography (e.g. Pittenger, 1976) and economics (e.g. Milne et al, 1980), although in practical applications they generally include some features of mixed models. In this section, some top-down models will be reviewed and it will be shown how variations between items may be introduced to affect the fractions. The simplest top-down model is the linear difference model:

\[ x_i(t+1) - x_i(t) = x_j(t+1) - x_j(t) = \frac{1}{N} \left[ x_+(t+1) - x_+(t) \right] \]  \hspace{1cm} (1)

where \( N \) is the number of items in the classification and \( x_+(t) \) is the level of the aggregate at time \( t \)

\[ x_+(t) = \sum_{i} x_i(t) \]. According to this model, the level change is the same for each item. The projection model that results is:

\[ x_i(t+1) = x_i(t) + \frac{1}{N} \left[ x_+(t+1) - x_+(t) \right] \]  \hspace{1cm} (2)
Another simple model is the ratio method. Each item is assumed to be a constant fraction of the aggregate:

\[ x_i(t) = p_i x_+(t), \quad \text{with} \quad \sum_{i} p_i = 1. \]  

The forecasts of the individual items are obtained by applying the time-independent fractions \( p_i \) to the forecasted aggregate. The ratio method implies that (i) the ratio of the levels of two items remains constant:

\[ \frac{x_i(t)}{x_j(t)} = \frac{x_i(t+1)}{x_j(t+1)}. \]  

(ii) each item changes at the same rate:

\[ \frac{x_i(t+1)}{x_i(t)} = \frac{x_j(t+1)}{x_j(t)} = \frac{x_+(t+1)}{x_+(t)}. \]  

Four remarks may be made here. First (3) represents an independence model (independence of fraction and time). Second (4) and (5) are equivalent to an odds ratio of one. Third equations (4) and (5) may be written as first-order differences of logarithms of levels. Let \( Z_i(t) = \ln x_i(t) \), then (5) becomes:

\[ Z_i(t+1) - Z_i(t) = Z_j(t+1) - Z_j(t). \]

Hence, the ratio method implies that the first-order differences of the transformed levels (log levels) are the same for each time series in the multiple series. Finally, the ratio method may be written as the following difference equation:

\[ x_i(t+1) = x_i(t) + p_i \left[ x_+(t+1) - x_+(t) \right]. \]  

The forecasts of \( x_+(t) \) completely determine the forecasts of \( x_i(t) \).
More complex top-down models may be distinguished on the basis of what they keep constant. The identification of an optimal top-down model may be viewed to be analogous to the usual differencing of a univariate time series. By appropriate transformation and differencing the multiple observations at a given time point and/or along the time series, constant relations between the component time series may be obtained. For instance, we may keep the change in growth ratio constant, instead of the growth ratio as in (5). If in addition the change is the same for all items, then the projection model may be derived as follows:

\[
\frac{x_i(t+1)}{x_i(t)} / \frac{x_i(t)}{x_i(t-1)} = \frac{x_+(t+1)}{x_+(t)} / \frac{x_+(t)}{x_+(t-1)} = r
\]

where \( r \) is the change in the aggregate growth ratio. The constancy of the change in the growth ratio of a particular item \( i \) is equivalent to a constant ratio of the growth ratio of item \( i \) and the overall (aggregate) growth ratio. Equation (7) is equivalent to

\[
\frac{x_i(t+1)}{x_i(t)} / \frac{x_i(t)}{x_i(t-1)} = \frac{x_+(t)}{x_+(t-1)} / \frac{x_+(t)}{x_+(t-1)} = 1
\]

which shows that the assumption of constant change in growth ratio implies that the growth ratio of \( i \) is equal to the aggregate growth ratio. Equation (7) may still be written in another way:

\[
\frac{x_i(t+1)}{x_+(t+1)} / \frac{x_i(t)}{x_+(t)} = 1
\]

which shows that the share of item \( i \) grows at a constant ratio. The different perspectives on the same constancy are helpful in interpreting projection models.

Expression (7) is equivalent to:

\[
\left[ \ln x_i(t+1) - \ln x_i(t) \right] - \left[ \ln x_i(t) - \ln x_i(t-1) \right] = \\
\left[ \ln x_+(t+1) - \ln x_+(t) \right] - \left[ \ln x_+(t) - \ln x_+(t-1) \right]
\]
which denotes equality of the second-order difference of the transformed variable. It is equal to:

\[ \ln x_1(t+1) - 2 \ln x_1(t) + \ln x_1(t-1) = \ln x_1'(t+1) - 2 \ln x_1'(t) + \ln x_1'(t-1). \]  \hspace{1cm} (8)

The constant-change-in-growth-ratio method may also be written as follows:

\[ x_1(t+1) / x_1(t) = q_1 \left[ x_1'(t+1) / x_1'(t) \right], \]

which leads to the following projection model:

\[ x_1(t+1) = q_1 \frac{x_1'(t)}{x_1(t)} \hspace{1cm} (9) \]

with \( \sum_i q_i x_1(t) = x_1(t) \).

In contrast to (6), (9) is a multiplicative model. Note that (9) is equivalent to the quadratic projection model (10) (quadratic in log):

\[ \ln x_1(t+1) = 2 \ln x_1(t) - \ln x_1(t-1) + \ln x_1'(t+1) - 2 \ln x_1'(t) + \ln x_1'(t-1). \]  \hspace{1cm} (10)

Several other top-down models may be imagined. A review is beyond the scope of this paper. Ter Heide (1981) presents an overview of distribution techniques for regional population projection.

The models presented above postulate equality of the n-th order differences (first and second) for all items in the classification and hence also for the aggregate time series. Equivalently, the ratio of the n-th order difference of two (transformed) individual item time series is equal to one. This equality is an unnecessary restriction. Ter Heide (1981, p. 12) for instance, considers a regional population projection model which implies that the ratio of the second-order difference of the population series of a particular region to that of the nation is equal to the fraction of the total (national) population that lives in the region in the base period. According to that model, the ratio of changes in regional and national population growth (in absolute terms) is equal to that of the regional and national populations.

The fixed ratio of n-th order differences may itself not be an adequate representation of reality. Some authors suggest exponential smoothing as an adjustment procedure. Thomopoulos (1980, p. 274) for instance applies a first-
order exponential smoothing to adjust the fractions (and hence the ratios of the untransformed variables). In general, let \( q_i(t) \) be the observed ratio of the \( n \)-th order differences of the time series of item \( i \) and the time series of the aggregate and let \( s_i(t) \) be the forecasted ratio of \( n \)-th order differences. Both differences are centered around \( t \). The first-order exponential smoothing model updates \( s_i(t) \) by adjusting the old estimate by some fraction \( \alpha \) of the forecast error (for the technique, see also Johnson and Montgomery, 1979, p. 32):

\[
s_{i}(t+1) = s_{i}(t) + \alpha [q_{i}(t) - s_{i}(t)],
\]

which after rearranging yields:

\[
s_{i}(t+1) = \alpha q_{i}(t) + (1 - \alpha) s_{i}(t).
\] (11)

The exponential smoothing constant \( \alpha \) is chosen to be close to 0 (e.g., \( \alpha = 0.1 \) or 0.2). Note that in this model the ratio between change in forecasted value and forecast error is fixed. The constant \( \alpha \) may also be chosen to minimize the sum of squared forecast errors.

5.2. **ER-models: state-space models**

5.2.1. **Model identification**

ER models describe the system dynamics in terms of the linkages between the system elements. The models may be written in state-space form. In this paper we consider the linear model:

\[
x(t+1) = P(t) x(t) + F(t) u(t) + C(t) w(t)
\] (12)

\[
y(t) = D(t) x(t) + v(t)
\] (13)

where \( x(t) \) is the state vector, \( u(t) \) the input vector of exogenous factors (elements in the environment of the system), \( w(t) \) is a vector of random shocks during the evolution of the process (process noise), \( y(t) \) is the output vector of measurements on \( x(t) \) and \( v(t) \) is a vector of random measurement errors. The process noise and the measurement noise are assumed to be independent normal distributed with zero mean. The coefficient matrices \( P, F, C \) and \( D \) are time-
varying. The time dependence will be discussed later in this section. For the moment we assume time-invariant coefficient matrices.

If the state of the system can be observed without error, then the systems dynamics can be described by (12) only. Equation (12) represents a first-order autoregressive process. Omitting the exogenous factors and assuming that the process noise represents the forecasting error (i.e., \( C = I \)), it may be written as follows (with constant coefficients):

\[
x(t+1) - P x(t) = w(t). \tag{14}
\]

The one-step-ahead forecast is given by \(^2\)

\[
\hat{x}(t+1) = P x(t) \tag{15}
\]

In the bivariate case, we have

\[
x_1(t+1) - p_{11} x_1(t) - p_{21} x_2(t) = w_1(t)
\]

\[
x_2(t+1) - p_{12} x_1(t) - p_{22} x_2(t) = w_2(t) \tag{16}
\]

If the backward shift operator \( B \) is introduced, then

\[
B x_1(t+1) = x_1(t) \tag{17}
\]

and (16) may be written as follows:

\[
\begin{bmatrix}
1 - p_{11} B & -p_{21} B \\
-p_{12} B & 1 - p_{22} B
\end{bmatrix}
\begin{bmatrix}
x_1(t+1) \\
x_2(t+1)
\end{bmatrix}
= \begin{bmatrix}
w_1(t) \\
w_2(t)
\end{bmatrix} \tag{18}
\]

Equation (18) may be written as:

\[
\begin{bmatrix}
I - P B
\end{bmatrix}
x(t+1) = w(t) \tag{19}
\]

Notice that (18) is equivalent to the expression of the bivariate linear model, given by Jenkins and Alavi (1981, p. 6, equation 2.10). There are a few differences between the usual state-space form and the more common form in time series analysis. The main difference is the backward shift operator notation adopted in time series analysis. Other differences are editorial; they are given in note 1 at the end of this paper. The equivalence of the state-space form and the usual multivariate autoregressive moving average form is important, since it allows a transition from one form to the other.
If $D$ is the identity matrix and the process noise is absent, then the state-space model is:

$$y(t+1) = P \cdot y(t) + F \cdot u(t) + \left[ v(t+1) - P \cdot v(t) \right], \quad (20)$$

which is a first-order autoregressive moving average model.

The state-space model is a general form. Also autoregressive moving average processes of order $(p, q)$ and of any dimension (number of individual time series in multiple series) may be transformed into a state space model. Consider the multivariate ARMA model $^3$:

$$\phi(B) \cdot z(t) = \Theta(B) \cdot a(t) \quad (21)$$

where $\phi(B)$ is the autoregressive operator

$$\phi(B) = 1 - \phi_1 B - \phi_2 B^2 - \phi_3 B^3 \ldots \phi_p B^p \quad (22)$$

and $\Theta(B)$ is the moving average operator

$$\Theta(B) = 1 - \theta_1 B - \theta_2 B^2 - \theta_3 B^3 \ldots \theta_q B^q \quad (23)$$

$\phi(B)$ and $\Theta(B)$ are known as transfer functions, and $z(t)$ is a stationary time series with mean zero and is obtained from the original series $x(t)$ by appropriate transformation and sufficient differencing

$$z_i(t) = \nabla^{d_i} x_i^{\{\lambda_i\}}(t),$$

with $x_i^{\{\lambda_i\}}(t)$ a non-linear transformation of the original time series, $\nabla$ the backward difference operator and $d_i$ is the order of differencing, needed to convert the original series $x_i(t)$ to a stationary time series.

$\phi(B)$ and $\Theta(B)$ are matrices with elements, respectively $\phi_{ij}(B)$ and $\theta_{ij}(B)$ with $i$ referring to a particular individual time series and $j$ to another time series to which $i$ is related. The diagonal elements of the operators are unity. The degrees of the polynomials (degrees of the backward shift operators) are respectively $p_{ij}$ and $q_{ij}$. In (22) and (23), $p$ and $q$ are equal to:

$p = \max (p_{ij})$ and $q = \max (q_{ij})$. The optimal values of $p_{ij}$, $q_{ij}$ and $d_i$ are determined from an analysis of past data, more specifically the analysis of the observed autocorrelation functions (model identification phase).
autocorrelation function is the principal tool used for this purpose. We would like to note here that differencing individual series of a multiple series involves interpretational (Hsiao, 1979) and some technical (Lütkepohl, 1982) problems, thus it is not so straightforward in multivariate case as in univariate case.

The multivariate ARIMA model is studied by Jenkins and Alavi (1981). Model (21) may be rewritten as follows:

\[ z(t) - \sum_{j=1}^{p} \phi_j z(t-j) = a(t) - \sum_{j=1}^{q} \theta_j a(t-j) \] (24)

\[ z(t) = \sum_{j=1}^{p} \phi_j z(t-j) + \sum_{j=0}^{q} \theta_j a(t-j). \] (24')

Expression (24') is a convenient formulation for forecasting purposes (see also Granger and Newbold, 1977, p. 24).

Let \( n = \max \{ p, q + 1 \} \). In order to transform the above \( n \)-th order difference equation into a state-space model of \( n \) first-order difference equations, we must define new variables and equations. The transformation is common in dynamic systems analysis. For instance, Kenkel (1974, pp. 295-299) shows the relation between the reduced form of an econometric model and the state-space form. Pindyck (1973, pp. 89-94) demonstrates how the structural form may be transformed into the state-space form (see also Chow, 1975 and Aoki, 1976). These transformations were used by Willekens (1976) in developing models for dynamic population distribution policy making. The key link between a higher-order difference equation and a system of first-order difference equations is a matrix, which in mathematical systems theory is known as the companion matrix. The same principles may be used to transform the conventional model (24) of a zero mean stationary autoregressive-moving average model into a state-space model, as was shown by Akaike (1974), Jenkins (1979, p. 115) and Priestley (1980, pp. 49-50).

The representation of the ARMA \( (p, q) \) form in state-space form is not unique. The discussion of the state-space form may be fixed by choosing the minimal (irreducible) form, however, still the state-space model of dimension \( n \) may be transformed by a similarity transformation using a non-singular arbitrary \( n \times n \) matrix \( (T, \text{say}) \) and transforming \( \mathbf{x}^*(t) = T\mathbf{x}(t) \) (Priestley, 1980). Transformation of a non-stationary time series with seasonal and trend components into a state-space form is studied by Akaike (1979, 1981), Kitagawa (1981) and De Beer (1983). Since stationarity models may be regarded as plausible assumptions to local behaviour (Anderson, 1977), the flexibility of the
state-space form for incorporating non-stationarity by time varying coefficients may be exploited for long-range forecasts where stationarity assumption becomes questionable (Parzen, 1982). Priestley (1980) proposes a very general state dependent model which is non-linear and non-stationary, that can conveniently be represented in state-space form and its estimation by using a sequential (Kalman filter type) algorithm. He also points to ways in which notions of mathematical systems theory may fruitfully be used in time series analysis (see also Jenkins, 1979, pp. 115-116). The transformation of conventional models of systems dynamics into the state-space model leads to many new perspectives for the investigation of dynamic systems. One perspective is the direct applicability of the Kalman filter technique (Mehra, 1979). Newbold states that: "The advantage of the state-space representation is that it readily allows through a recursive algorithm, the Kalman filter (....), the estimation of parameters and the prediction of future values" (Newbold, 1981, p. 61). The Kalman filter technique may also enhance the multivariate time series analysis and forecasting with incomplete data (Ansley and Kohn, 1983).

The state-space model arose in many disciplines as a unified framework for the analysis and forecasting of dynamic systems. The transformation of conventional models of systems dynamics into the state-space model may help to clarify issues that would otherwise remain unsolved. The main advantage of the state-space model is, however, its simple structure. The properties of the model have extensively been studied in the mathematical systems theory and the optimal control literature (see e.g. Wolovich, 1974). In the forecasting literature, the model is not viewed unanimously as an advance over other available methods. Jenkins (1982, p. 12) for instance rejects the method because it shall be based on continuous changes in the parameter values (each time an observation comes at hand), for which no statistical framework exists. Jenkins and Alavi (1981, p. 43) propose to generate multivariate forecasts from the difference equation form of (21). The forecasts of $z(t+1)$ are generated recursively with $l = 1, 2, \ldots$ by replacing future values of $z(t+1)$ by their forecasts made at time $t$, $\hat{z}_t(l)$ and future values by the random residual series $\alpha(t+1)$ by 0. Other authors are more sceptical about the potential of model (21). Newbold (1979, p. 69) for instance recommends a marriage between time-series methods and behavioral model-building techniques to tackle the problems of multivariate time-series analysis.

5.2.2. Parameter estimation

The application of ER-models to multidimensional forecasting requires a knowledge of the coefficient matrices. If the elements can be measured without
error and if the linkages between the system elements and between the system and the environment can directly be measured, then the matrices $P$ and $F$ can easily be determined. The appropriate framework is the accounting framework. If the linkages cannot be observed, they might be estimated from the data. The appropriate framework in this case is the statistical framework. Both frameworks will be discussed below.

a. ER-models in an accounting framework.

The data obtained by an ER-observation plan may be arranged in an account (Table 2). If we assume no measurement error and no process noise, then the ER-model may be derived from the account in three steps. In the first step, accounting equations relating $x(t+1)$ to $x(t)$ are derived. For instance:

$$x_i(t+1) = a_{ii}(t) + \sum_{k \neq i} a_{ki}(t) + a_{ei}(t)$$

$$x_i(t) = a_{ii}(t) + \sum_{j \neq i} a_{ij}(t) + a_{ie}(t)$$

$$x_i(t+1) = x_i(t) - \sum_{j \neq i} a_{ij}(t) + \sum_{k \neq i} a_{ki}(t) - a_{ie}(t) + a_{ei}(t)$$

In the second step, transition probabilities are calculated. In the simplest case the transition probability $p_{ij}(t) = a_{ij}(t) / x_i(t)$. Sometimes, the observation plan requires some intermediate steps, since $p_{ij}(t)$ are improper probabilities, which are not necessarily less than one. The probabilities constitute the elements of the matrix $P$. Since transition probabilities cannot be calculated for linkages originating in the environment of the system, these linkages are treated as elements of an input vector. The third step is the design of the model. The model is:

$$x_i(t+1) = \left[ 1 - \sum_{j \neq i} p_{ij}(t) - p_{ie}(t) \right] x_i(t) + \sum_{k \neq i} p_{ki}(t) x_k(t) + a_{ei}(t).$$

The model, written for all $i$, may yield the following matrix expression:

$$x(t+1) = P(t)x(t) + a_e(t),$$

where the diagonal elements of $P(t)$ are $p_{ii}(t) = 1 - \sum_{j \neq i} p_{ij}(t) - p_{ie}(t)$.
b. ER-models in a statistical framework.

If the elements of the coefficient matrices cannot be calculated directly from observations, they must be estimated. The data availability may be limited to a time series of past state vectors \( x(t) \). Some partial and/or aggregate information on the linkages may, however, also be available.

(i) Estimation of coefficient matrices from time series of state vectors.

In the literature, two approaches are presented to estimate the coefficient matrices from vector time series. The first approach is suited for first-order autoregressive models without exogenous factors (Markov chains). The estimation problem is defined as a problem of estimating the terms of the transition matrix of a Markov chain. If least-square estimates ought to be obtained, the problem is a quadratic programming problem with the estimates restricted to be non-negative. The approach is extensively discussed by Lee et al (1970).

The second approach is a generalization of estimation methods for univariate time series to multivariate time series. The problem is to find values for \( \phi_{ij} \) and \( \theta_{ij} \). Maximum likelihood methods which are equivalent to iterative reweighted least squares procedures are discussed by Jenkins and Alavi (1981, pp. 32-34).

(ii) Estimation of coefficient matrices from incomplete information on the linkages.

If incomplete, frequently aggregate, information on the linkages is available, the coefficients associated with missing linkages are estimated by maximizing the likelihood of their occurrences. The approach is equivalent to multipropotionally adjusting initial guesses of linkages to satisfy the known information. A review of techniques is given by Willekens (1982a). This estimation problem is frequently encountered in demographic models and in input-output models of economic change. In the latter application, the linkages between the sectors of the economy at a particular point in time \( t \) are estimated by adjusting a historical input-output table to the national economic account for \( t \). The approach is a special case of the procedure for forecasting structural change presented in the next section.
6. **STEP 5: SYSTEMS FORECASTING WITH VARIABLE COEFFICIENT MATRICES**

Multidimensional forecasting is the forecasting of a set of interrelated elements. Changes in the elements occur because of autonomous changes and because of the existing linkages and changes therein. In the previous section, the state-space model was selected as an appropriate forecasting model because it emphasizes the mutual dependence of elements of a whole, and hence of the individual time series. The linkages between the elements are represented by the off-diagonal elements of the coefficient matrices. It was assumed that the coefficient matrices are time-invariant, which implies that the relative linkages (e.g., transition probabilities) do not change over time and the levels or absolute values of the linkages change only because of variations in the levels of elements.

In this section we allow the coefficient matrices, i.e., the parameters of the forecasting model, to change. The focus will be on the \( P \)-matrix of transition probabilities; the arguments may equally well be applied to the other coefficient matrices. The topic of coefficient change received relatively little attention in the literature. Harrison and Stevens (1976) and Young (1978) developed algorithms for estimating time-dependent coefficients of linear systems, in which the coefficients may vary freely over time. In the Priestley (1980) model, the changes in the coefficients are determined by changes in the state variable. The Kalman filter has been proposed as an appropriate technique for the estimation of time-dependent coefficients (see also Mehra, 1979). Jenkins (1982, p. 12) questioned the usefulness of the Kalman filter. His main argument was that coefficients tend to change slowly. One could add that the coefficient changes are not independent. Some coefficients may change without their relation to other coefficients being affected. The fundamental issue is therefore the change in the relation between coefficients and hence the change in the relatedness of linkages between system elements. Such changes were denoted earlier in the paper as structural change. Structural change is the focus of this section.

The present section is exploratory in nature. At its origin is the speculation that recent developments in generalized linear models (GLM) may be of use to model time-varying transition matrices. The main advantage of these models, which resemble the analysis-of-variance paradigm, is that they decompose transition matrices into component effects, the time-dependence of which may be studied separately. The size of the problem may therefore substantially be reduced. In addition, each component effect is quantified and outliers can easily be identified. In this paper we focus on a particular type of GLM, namely the log-linear model.
The log-linear model of a two-way table $\mathbf{P}$ is

$$p_{ij} = w \cdot w^A_i \cdot w^B_j \cdot w^{AB}_{ij}$$

or

$$\ln p_{ij} = u + u^A_i + u^B_j + u^{AB}_{ij},$$

with

$$\pi w^A_i = \pi w^B_j = \pi w^{AB}_{ij} = \pi w^{AB}_{ij} = 1$$

and

$$\sum u^A_i = \sum u^B_j = \sum u^{AB}_{ij} = \sum u^{AB}_{ij} = 0.$$  

$w(u)$ is the overall effect, $w^A_i$ ($u^A_i$) is the row effect, $w^B_j$ ($u^B_j$) is the column effect and $w^{AB}_{ij}$ ($u^{AB}_{ij}$) is the interaction effect.

A review of generalized linear models is beyond the scope of this paper. The log-linear model, which is the form of GLM used in this paper, is discussed by Bishop, Fienberg and Holland (1975) among others. Very few authors have used the log-linear model for forecasting. Some attempts were presented by Clogg (1979, chapter 7).

Let a time series of transition matrices $\mathbf{P}(t)$ be given ($t=0, 1, 2, \ldots, T$). The matrices may be arranged in a three-dimensional table of dimension $N \times N \times T$, with $N$ being the number of elements in the system and hence the number of elements in the state vector and $T$ is the number of time points in the series. Let $\ast \mathbf{P}$ be an appropriate transformation of $\mathbf{P}$. In this section we consider the logarithm of probabilities. If the system is closed and only redistribution among the elements is allowed, then the column totals of the transition matrix are equal to one and the matrix is a Markov matrix.

The problem of forecasting structural change is a problem of updating and forecasting the elements of $\mathbf{P}$ or $\ast \mathbf{P}$ and of relations between the elements (e.g. odds ratios, elasticities). In this section we focus on updating the transition matrix of the base period $\mathbf{P}(0)$. The estimation of the transition matrix at time $t=1$ from $\mathbf{P}(0)$ depends on the information available on $\mathbf{P}(1)$. If no information is available the content of $\mathbf{P}(1)$ cannot be forecasted at all and we may assume constancy, i.e. $\hat{\mathbf{P}}(1) = \mathbf{P}(0)$. If the marginal totals of $\mathbf{P}(1)$ or of the gross flow matrix underlying $\mathbf{P}(1)$ are known, we may update $\mathbf{P}(0)$ to satisfy the given marginal totals. The updating equation is:

$$\hat{p}_{ij}(1) = p_{ij}(0) \cdot w^A_{0i} \cdot w^B_{0j},$$

where $w^A_{0i}$ and $w^B_{0j}$ are biproportional adjustment factors, which are estimated from
P(0) and the new marginal totals and which may be scaled to satisfy the normalization restriction \( \prod_i w_{i0} = \prod_j w_{j0} = 1 \). The factor \( w_{i0}^A \) is associated with row i, and \( w_{j0}^B \) is associated with column j. The matrix \( \hat{P}(1) \) is biproportional to \( P(0) \). The biproportional adjustment method is a popular updating technique in input-output analysis (Bacharach, 1970). The method is also frequently used, although in different forms, in transportation science and regional science to update spatial transaction matrices (Willekens, 1983). Forecasts of the marginal totals may be derived from time-series or econometric forecasting models or from any other forecasting technique.

Equation (30) may be written in matrix form:

\[
\hat{P}(1) = w_o^A P(0) w_o^B ,
\]

(31)

where \( w_o^A \) and \( w_o^B \) are diagonal matrices.

Equation (30) may also be written as follows:

\[
\ln \hat{p}_{ij}(1) = \ln \hat{p}_{ij}(0) + \ln w_{i0}^A + \ln w_{j0}^B ,
\]

or

\[
\hat{p}_{ij}(1) = \hat{p}_{ij}(0) + u_{i0}^A + u_{j0}^B ,
\]

where \( u_{i0} = \ln w_{i0}^A \), \( u_{i0}^A = \ln w_{i0}^A \) and \( u_{j0}^B = \ln w_{j0}^B \). The w- and u-coefficients are parameters of the unsaturated log-linear model for the matrix of growth ratios:

\[
R_{ij}(0, 1) = \frac{\hat{p}_{ij}(1)}{p_{ij}(0)} .
\]

Model (30) postulates the absence of interaction effects in the matrix \( R(0, 1) \).

\[
\hat{R}_{ij}(0, 1) = \frac{\hat{p}_{ij}(1)}{p_{ij}(0)} = \frac{w_{i0}^A w_{j0}^B}{w_{i0}^A w_{j0}^B} ,
\]

(32)

where \( r_o = \frac{w_{i0}}{w_{i0}^A} \), \( r_{j0} = \frac{w_{j0}}{w_{j0}^B} \), \( r_{i0} = \frac{w_{i0}^A}{w_{i0}} \) and \( r_{j0} = \frac{w_{j0}^B}{w_{j0}} \). This can easily be seen by writing the log-linear model for \( P(0) \) and \( P(1) \):

\[
\frac{\hat{p}_{ij}(1)}{p_{ij}(0)} = \frac{w_{i1}^A w_{j1}^B w_{ij1}^{AB}}{w_{i0}^A w_{j0}^B w_{ij0}^{AB}} ,
\]

with \( w_{ij1}^{AB} = w_{ij0}^{AB} \), as will be shown below. It can be shown that this absence of
interaction effects implies that \( \hat{P}(1) \) exhibit the same interaction effects as \( P(0) \). Willekens (1983) shows that the biproportional adjustment procedure transmits the interaction effects of the known historical matrix to the matrix to be estimated. The main effects are determined by the growth ratio of the row and column totals (ratio of new to old marginal totals). In general, the log-linear model parameters that are adjacent to the prior information on the matrix to be estimated are adjusted by the updating procedure. The other parameters remain constant. For instance, if only the column totals of the matrix to be estimated are given, then the row effects and the interaction effects are "borrowed" from the historical matrix and are therefore preserved in the updating.

We now turn to the estimation of the transition matrix at time \( t = 2 \). The matrix \( P(2) \) may depend on \( P(1) \) and \( P(0) \). The updating model may take different forms depending on the assumptions about what remains constant and what prior information on the forecast is given. The forecasting of \( \hat{P}(2) \) from \( P(1) \) is analogous to the forecasting of \( \hat{P}(1) \) from \( P(0) \). The assumption of constant interaction effects may be a workable assumption. In practice, however, \( P(1), \hat{P}(1) \) and \( P(0) \) should be used to obtain \( \hat{P}(2) \). Various approaches are possible to generate the forecast \( \hat{P}(2) \).

If no information on \( p_{ij}(2) \) is available, then the transition probabilities may be forecasted by the classical single exponential smoothing method. The forecasting equation is:

\[
\hat{p}_{ij}(2) = \hat{p}_{ij}(1) + \alpha \left[ p_{ij}(1) - \hat{p}_{ij}(1) \right],
\]

where \( \hat{p}_{ij}(1) \) is the estimate of \( p_{ij}(1) \), obtained by forecasting from \( p_{ij}(0) \) and \( \alpha \) is the smoothing constant. In this model the forecasts of the transition probabilities are adjusted on the basis of previous forecast errors. Equation (33) may be written as:

\[
\hat{p}_{ij}(2) = \alpha \cdot p_{ij}(1) + (1 - \alpha) \cdot \hat{p}_{ij}(1),
\]

which is similar to (1.1). The smoothing constant \( \alpha \) can be chosen to minimize the sum of squared forecast errors:

\[
\sum_{i,j} \left[ p_{ij}(2) - \hat{p}_{ij}(2) \right]^2.
\]

Equation (34) may be written as follows:

\[
\frac{\hat{p}_{ij}(2) - \hat{p}_{ij}(1)}{p_{ij}(1) - \hat{p}_{ij}(1)} = \alpha
\]

(35).
If both the numerator and the denominator are divided by \( \hat{p}_{ij}(1) \), \( \alpha \) may be interpreted as an elasticity. It is the percentage by which \( \hat{p}_{ij}(1) \) is changed (adjusted) because of a one percent forecast error. If \( \alpha = 0 \), the estimated transition probability is assumed to remain constant. If \( \alpha = 1 \), then \( \hat{p}_{ij}(2) = p_{ij}(1) \) and the one step ahead forecast of the transition probability is the observed probability.

The classical exponential smoothing model is widely used. It represents, however, only one way to adjust the forecasts on the basis of past errors. Other smoothing methods may be developed. Writing equation (35) as

\[
\frac{\hat{p}_{ij}(2) / \hat{p}_{ij}(1) - 1}{p_{ij}(1) / \hat{p}_{ij}(1) - 1} = \alpha,
\]

suggests the following method

\[
\frac{\hat{p}_{ij}(2) / \hat{p}_{ij}(1)}{p_{ij}(1) / \hat{p}_{ij}(1)} = \beta. \tag{36}
\]

Model (36) updates the growth ratio of \( \hat{p}_{ij} \) on the basis of the forecast error. An overestimation of \( p_{ij}(1) \) leads to a reduction in the growth ratio \( \hat{p}_{ij}(2) / \hat{p}_{ij}(1) \). The new growth ratio is \( \beta \cdot p_{ij}(1) / \hat{p}_{ij}(1) \). The ratio \( \beta \) may be chosen to minimize the forecast error. It may be obtained by fitting a linear regression model with intercept zero through the time series \( p_{ij}(t) \), \( t = 0, 1, 2, \ldots \) (all \( i \) and \( j \)). In the absence of forecast error, the growth ratio remains at a constant level \( \beta \), which represents a simple extrapolation. This method (36) is therefore suited for the adjustment of forecasts generated by a constant growth rate model. The adjustment method (36) implies the following forecasting equation:

\[
\hat{p}_{ij}(2) = \beta \cdot \hat{p}_{ij}(1) \left( \frac{p_{ij}(1)}{\hat{p}_{ij}(1)} \right), \tag{37}
\]

or

\[
\ln \hat{p}_{ij}(2) = \ln \hat{p}_{ij}(1) + \left[ \ln p_{ij}(1) - \ln \hat{p}_{ij}(1) \right] + \ln \beta.
\]

Equation (37) reduces to

\[
\hat{p}_{ij}(2) = \beta \cdot p_{ij}(1). \tag{38}
\]
If $\beta = 1$, the one step ahead forecast is fully adjusted to the observed probability. A comparison of (38) to (30) reveals that $\beta$ may be thought of as an overall effect and that the row, column and interaction effects of $P(1)$ are transmitted to $\hat{P}(2)$. The method (37) may therefore be referred to as an updating method which changes the overall effect only. The change in the overall effect is determined by past forecast errors.

The forecasting model (37) may be extended by allowing $\beta$ to vary with row $i$, column $j$ or cell $(i, j)$. If the smoothing constant changes with row $i$ and column $j$, then (36) may be written as

$$\hat{p}_{ij}(2) = r_i s_j \frac{p_{ij}(1)}{\hat{p}_{ij}(1)} \tag{39}$$

with $r_i$ and $s_j$ row and column smoothing constants, chosen as to minimize the forecast errors. We may rewrite the expression as the following forecasting model:

$$\hat{p}_{ij}(2) = r_i s_j p_{ij}(1). \tag{40}$$

The double extension of the exponential smoothing model to (i) a multiplicative model (or additive model in transformed variables) and (ii) a row and column specific smoothing constant leads to a biproportional adjustment model. The model was already presented in (30). Equation (40) may be written in the form of (30) by introducing a scalar $b$ to allow a proper scaling of the $r_i$ and $s_j$ coefficients: $\Pi i r_i = \Pi j s_j = 1$.

There is, however, an important difference between (30) and (40). In (30) the biproportional adjustment factors are estimated from the priorly known marginal totals of the new transition matrix. In (40) the factors are obtained by minimizing the forecast error

$$\sum_{i,j} \left[ p_{ij}(2) - \hat{p}_{ij}(2) \right]^2.$$

Because of the relation between the biproportional adjustment method and the log-linear model, the log-linear model provides a unified framework for multidimensional forecasting on the basis of past data, forecast errors and prior information on forecasts. Conventional forecasting methods, such as the exponential smoothing method, fit logically in this framework.

The combination of prior information on forecasts and on forecast errors will now be demonstrated. Suppose that we know $P(1)$ and the marginal totals of $P(2)$. The
forecasting model is the biproportional adjustment model. The w-coefficients may be estimated from the marginal totals. Recall that the interaction effects of P(2) are those of P(1). The forecast may be improved by applying the extended exponential smoothing model to the interaction effects. By doing so, it should be kept in mind that any change in interaction effects influences the main effects if given marginal totals of \( \hat{P}(2) \) must be satisfied. It follows from the marginality condition in hierarchical log-linear models.

The exponential smoothing model (33) may also be extended in a different way. Instead of smoothing the probabilities \( p_{ij} \), we may smooth the transformed probabilities \( \ln p_{ij} \). The transformation ensures positive elements for the forecasted probability matrix. The forecasting equation becomes:

\[
\ln \hat{p}_{ij}(2) = \ln \hat{p}_{ij}(1) + \gamma \left[ \ln p_{ij}(1) - \ln \hat{p}_{ij}(1) \right]
\]  

(41)

The equation may be written as follows:

\[
\frac{\hat{p}_{ij}(2)}{\hat{p}_{ij}(1)} = \left[ \frac{p_{ij}(1)}{\hat{p}_{ij}(1)} \right]^\gamma
\]

and \( \hat{p}_{ij}(2) = [p_{ij}(1)]^\gamma \left[ \hat{p}_{ij}(1) \right]^{1-\gamma} \)  

(42)

In this updating method, the forecasted transition probability at time \( t=2 \) is a weighted geometric mean of the observed and the forecasted probability at time \( t=1 \). Notice that (42) resembles the Cobb Douglas production function with inputs the observed probability and the estimated probability. If \( \gamma = 0 \), then \( \hat{p}_{ij}(2) = \hat{p}_{ij}(1) \) and no adjustment takes place. If \( \gamma = 1 \), then \( \hat{p}_{ij}(2) = p_{ij}(1) \) and the one step ahead forecast is equal to the currently observed probability.

Notice the difference between (42) and (34). In (34) the one step ahead forecast is a linear combination of the observed and forecasted probability. Equation (42) denotes a geometric combination, or a linear combination in the transformed variables:

\[
\ln \hat{p}_{ij}(2) = \gamma \ln p_{ij}(1) + (1-\gamma) \ln \hat{p}_{ij}(1).
\]

The log transformation of the probabilities ensures positive elements of the probability matrices. The probabilities must, however, also lie between zero and unity. This may be ensured by a logit transformation. The forecasting equation becomes:
\[
\ln \frac{\hat{p}_{ij}(2)}{1 - \hat{p}_{ij}(2)} = \ln \frac{\hat{p}_{ij}(1)}{1 - \hat{p}_{ij}(1)} + \gamma \left[ \ln \frac{p_{ij}(1)}{1 - p_{ij}(1)} - \ln \frac{\hat{p}_{ij}(1)}{1 - \hat{p}_{ij}(1)} \right] \tag{43}
\]

which implies
\[
\frac{\hat{p}_{ij}(2)}{1 - \hat{p}_{ij}(2)} = \left[ \frac{p_{ij}(1)}{1 - p_{ij}(1)} \right]^{\gamma} \left[ \frac{\hat{p}_{ij}(1)}{1 - \hat{p}_{ij}(1)} \right]^{1 - \gamma},
\]

where \( \gamma \) is selected to minimize the forecast error subject to the constraint that \( \sum_j \hat{p}_{ij}(2) = 1. \)
7. CONCLUSION

Multidimensional forecasting is the forecasting of interrelated elements, which may relate to different variables, to items of a group or to categories of a classified variable. In this paper a series of procedures is described which may assist the forecaster. The bottom-line of the exposition is that to forecast interrelated elements we have to focus on the linkages. This can best be done by adopting a systems analytic view on forecasting. To forecast interrelated elements, we view the elements as parts of a system. A system is more than a collection of elements. It exhibits a structure, which originates from particular relations among the elements and which is as important to understand the system as the elements. To investigate the dynamics of a system, we may use notions and models, developed in mathematical systems theory. It has been demonstrated in this paper and elsewhere by several authors that mathematical systems theory provides the analytical framework for the analysis, forecasting and control of dynamic systems.

The identification of the linkages is the first step in multidimensional forecasting. Ideally, the existence of linkages between elements leads to events, which generate transactions or flows and which in turn cause elements to change due to increments and decrements. The linkages of interest are frequently not observable, either due to the inherent character of the phenomenon or due to the complexities of linkages (such as the relation between inflation and unemployment). What may be measured are symptoms of the existence of linkages, such as differences in the state of the system at two consecutive points in time. A vector time series is an illustration of such an indirect measurement. Statistical techniques of multivariate time series analysis may be applied to infer the linkages. If the linkages can be observed, such as in demographic systems, the multidimensional forecasting process is simplified. The observability of linkages leads to two groups of observation plans. In E-observation plans only the system elements can be measured at different points in time; in ER-observation both the elements and the linkages can be known. The data gathered may be arranged in an account. We strongly recommend the adoption of an accounting framework to organize and prepare the data for multidimensional forecasting. The account not only gives an overview of the available data, but also provides an instrument for checking the internal consistency of the data and a framework for estimating missing data. ER-accounts have the additional advantage that the parameters of the state-space forecasting model may be derived directly from the account.

With each group of observation plans is associated a particular analysis and modelling perspective. The analysis of past and present dynamics of the system is a
prerequisite for good forecasting. An important aspect of the systems analysis is the quantification of the association between individual time series. A common measure of association in time series analysis is the cross-correlation function. Other measures presented in the paper to quantify relations between linkages are the odds ratio and the elasticity. It is shown that there exist different ways of writing these measures, each of which highlights a particular feature of the association (relatedness).

The systems analysis is the basis for systems modelling. A main conclusion of the paper is that the selection of a type of forecasting model should be based on the analysis of the relatedness of system elements and of linkages between elements. If the identification, measurement and analysis of the dynamic system is done well, the development of the forecasting model becomes relatively straightforward. We distinguish between two types of forecasting models. E-models forecast the system on the basis of knowledge of the state of the system at consecutive points in time only. ER-models forecast the system on the basis of both the state of the elements and the linkages. The linkages do not have to be directly observable to derive an ER-model; they may also be estimated either from a time series of states of the elements or from incomplete but relevant information on the linkages. In E-models only system elements are considered. Unless the individual elements are independent and their time path may be described by a set of unrelated univariate time series, changes in an element depend on changes in the other elements and/or in the aggregate. The dependence of changes of individual elements may be modelled by distribution functions. In this paper we reviewed several methods to allocate the forecasted aggregate among the individual elements. The relation with exponential smoothing models was shown. ER-models, which describe the systems dynamics in terms of the linkages between the system elements, may effectively be written in the form of state-space models. The state-space model has arisen in many disciplines as a unified framework for the analysis, forecasting and control of dynamic systems. It serves also as a unified approach to modelling for multidimensional forecasting. The conventional models for dynamic analysis, such as the multivariate ARMA models, may be transformed into the state-space format. The transformation is based on a simple principle, namely that an n-th order difference equation may be written as a system of n first-order difference equations. By transforming forecasting models into the state-space format, the arsenal of notions and techniques of mathematical systems theory becomes directly applicable for forecasting.

A tricky problem in multidimensional forecasting is the time-dependence of the coefficient matrices. For instance, the relative linkages, represented by transition probabilities, do not remain constant in time. The changes are, however, generally
not random, but follow some patterns. To describe the patterns of change of the coefficients of the state-space model, we suggest a generalized linear model (GLM). Generalized linear models decompose the coefficient matrices into component effects, the time-dependence of which may be studied separately. The size of the problem may therefore substantially be reduced. It was shown that the problem of updating and predicting the transition probability matrix of the state-space model is clarified and simplified by the adoption of a GLM. It was also shown that the exponential smoothing of the probabilities fits in this framework and that the GLM-approach leads to some very interesting and promising extensions of the smoothing methods. For instance, the log-linear model, a particular type of GLM, provides a framework to predict transition probability matrices from past data, forecast errors and prior information on forecasts.
NOTES

1) Defining structural change is not an easy task. Different authors use different definitions. In this paper, we refer to structural change if the relation between the parameters of the forecasting model changes. The change in model structure is therefore not a necessary condition for structural change.

2) In the time series literature, the one-year-ahead forecast on the basis of \( x(t) \) would be denoted by \( \hat{x}(t) \) or \( \hat{x}_t(1) \). We denote the forecast by \( \hat{x}(t+1) \). In the time series literature, the coefficients of \( P \) are referred to as weight functions (Jenkins and Alavi, 1981, p. 4). The forecast error \( \hat{x}(t) - x(t+1) \) is denoted by \( w(t+1) \), instead of by \( w(t) \) as in (14).

3) This model was first studied by Quenouille (1957).
REFERENCES

Akaike, H.  

Akaike, H.  

Akaike, H.  

Altham, P.M.E.  

Anderson, O.D.  

Ansley, C.F. and R. Kohn,  
1983 - Exact likelihood of vector autoregressive-moving average process with missing or aggregated data, Biometrika, 70, pp. 275-278.

Aoki, M.  
1976 - Optimal control and system theory in dynamic economic analysis, Amsterdam, North Holland.

Armstrong, J.S.  

Bacharach, M.  

Bartholomew, D.G.  

Batey, P.W.J. and M. Madden  

Baydar, N.  

Beer, J. de  
Bishop, Y.M., S.E. Fienberg and P.W. Holland

Blin, J.M., E.A. Stohr and B. Bagamery

Box, G.E.P. and G.M. Jenkins

Brown, R.G.

Chow, G.C.

Clogg, C.C.

Conway, Jr., R.S.

Fuiles, R.

Granger, C.W.J. and P. Newbold

Harrison, P.J. and C.F. Stevens

Heckman, J.J. and B. Singer

Heide, H. ter
1981 - Demographic distribution formulas, paper contributed to the 19th General Population Conference, Manila.

Hogarth, R.M. and S. Makridakis

Hsiao, C.

Jenkins, G.M.
Jenkins, G.M.

Jenkins, G.M. and A.S. Alavi

Johnson, L.A. and D.C. Montgomery

Juster, F.T. and K.C. Land (eds.)

Kenkel, J.L.

Keyfitz, N.

Kitagawa, G.

Land, K. and A. Rogers (eds.)

Lee, T.C., G.G. Judge and A. Zellner,
1970 - Estimating the parameters of the Markov probability model from aggregate time series data, Amsterdam, North Holland.

Leontief, W.
1941 - The structure of the American economy, 1919-1939, New York, Oxford University Press.

Lütkepohl, A.

Makridakis, S.

Makridakis, S.

Makridakis, S. and S.C. Wheelwright
Mehra, R.K.

Milne, W.J., F.G. Adams and N.J. Glickman

Newbold, P.

Newbold, P.

Parzen, E.

Pindyck, R.S.
1973 - Optimal planning for economic stabilization: the application of control theory to stabilization policy, Amsterdam, North Holland.

Pittenger, D.B.

Priestley, M.B.

Quenouille, M.H.
1957 - The analysis of multiple time series, London, Griffin.

Rees, P.H. and A.G. Wilson

Reinsel, G.
1983 - Some results on multivariate autoregressive index models, Biometrika, 70, pp. 145-156.

Rogers, A.
1975 - Introduction to multiregional mathematical demography, New York, Wiley.

Round, J.I.
1978 - On estimating trade flows in interregional input-output models, Regional Science and Urban Economics, 8, pp. 289 - 302

Steece, B.M. and S.D. Wood
Thomopoulos, N.T.

United Nations
1975 - Towards a system of social and demographic statistics, New York, UN Studies in Methods, Series F., No. 18.

Willekens, F.J.

Willekens, F.J.

Willekens, F.J.

Willekens, F.J.

Wolovich, W.

Young, P.C.
LIST OF WORKING PAPERS


* Out of print


