1

Introduction

1.1 Objectives of Analyzing Multiple Time Series

In making choices between alternative courses of action, decision makers at all structural levels often need predictions of economic variables. If time series observations are available for a variable of interest and the data from the past contain information about the future development of a variable, it is plausible to use as forecast some function of the data collected in the past. For instance, in forecasting the monthly unemployment rate, from past experience a forecaster may know that in some country or region a high unemployment rate in one month tends to be followed by a high rate in the next month. In other words, the rate changes only gradually. Assuming that the tendency prevails in future periods, forecasts can be based on current and past data.

Formally, this approach to forecasting may be expressed as follows. Let y_t denote the value of the variable of interest in period t. Then a forecast for period $T+h$, made at the end of period T, may have the form

$$\widehat{y}_{T+h} = f(y_T, y_{T-1}, \dots), \tag{1.1.1}$$

where $f(\cdot)$ denotes some suitable function of the past observations y_T, y_{T-1}, ... For the moment it is left open how many past observations enter into the forecast. One major goal of univariate time series analysis is to specify sensible forms of functions $f(\cdot)$. In many applications, linear functions have been used so that, for example,

$$\widehat{y}_{T+h} = \nu + \alpha_1 y_T + \alpha_2 y_{T-1} + \cdots .$$

In dealing with economic variables, often the value of one variable is not only related to its predecessors in time but, in addition, it depends on past values of other variables. For instance, household consumption expenditures may depend on variables such as income, interest rates, and investment expenditures. If all these variables are related to the consumption expenditures it makes sense to use their possible additional information content in forecasting consumption expenditures. In other words, denoting the related variables by $y_{1t}, y_{2t}, \dots, y_{Kt}$, the forecast of $y_{1,T+h}$ at the end of period T may be of the form

$$\widehat{y}_{1,T+h} = f_1(y_{1,T}, y_{2,T}, \dots, y_{K,T}, y_{1,T-1}, y_{2,T-1}, \dots, y_{K,T-1}, y_{1,T-2}, \dots).$$

Similarly, a forecast for the second variable may be based on past values of all variables in the system. More generally, a forecast of the k-th variable may be expressed as

$$\widehat{y}_{k,T+h} = f_k(y_{1,T}, \dots, y_{K,T}, y_{1,T-1}, \dots, y_{K,T-1}, \dots). \tag{1.1.2}$$

A set of time series y_{kt}, $k = 1, \dots, K$, $t = 1, \dots, T$, is called a *multiple time series* and the previous formula expresses the forecast $\widehat{y}_{k,T+h}$ as a function of a multiple time series. In analogy with the univariate case, it is one major objective of multiple time series analysis to determine suitable functions $f_1, \dots, f_K$ that may be used to obtain forecasts with good properties for the variables of the system.

It is also often of interest to learn about the dynamic interrelationships between a number of variables. For instance, in a system consisting of investment, income, and consumption one may want to know about the likely impact of a change in income. What will be the present and future implications of such an event for consumption and for investment? Under what conditions can the effect of an increase in income be isolated and traced through the system? Alternatively, given a particular subject matter theory, is it consistent with the relations implied by a multiple time series model which is developed with the help of statistical tools? These and other questions regarding the structure of the relationships between the variables involved are occasionally investigated in the context of multiple time series analysis. Thus, obtaining insight into the dynamic structure of a system is a further objective of multiple time series analysis.

1.2 Some Basics

In the following chapters, we will regard the values that a particular economic variable has assumed in a specific period as realizations of random variables. A time series will be assumed to be generated by a stochastic process. Although the reader is assumed to be familiar with these terms, it may be useful to briefly review some of the basic definitions and expressions at this point, in order to make the underlying concepts precise.

Let $(\Omega, \mathcal{F}, \Pr)$ be a *probability space*, where Ω is the set of all elementary events (sample space), $\mathcal{F}$ is a sigma-algebra of events or subsets of Ω and Pr is a probability measure defined on $\mathcal{F}$. A *random variable* y is a real valued function defined on Ω such that for each real number c, $A_c = \{\omega \in \Omega | y(\omega) \leq$

$c\} \in \mathcal{F}$. In other words, A_c is an event for which the probability is defined in terms of Pr. The function $F : \mathbb{R} \to [0, 1]$, defined by $F(c) = \Pr(A_c)$, is the distribution function of y.

A K-dimensional *random vector* or a K-dimensional *vector of random variables* is a function y from Ω into the K-dimensional Euclidean space $\mathbb{R}^K$, that is, y maps $\omega \in \Omega$ on $y(\omega) = (y_1(\omega), \dots, y_K(\omega))'$ such that for each $c = (c_1, \dots, c_K)' \in \mathbb{R}^K$,

$$A_c = \{\omega | y_1(\omega) \le c_1, \dots, y_K(\omega) \le c_K\} \in \mathcal{F}.$$

The function $F : \mathbb{R}^K \to [0, 1]$ defined by $F(c) = \Pr(A_c)$ is the joint *distribution function* of y.

Suppose Z is some index set with at most countably many elements like, for instance, the set of all integers or all positive integers. A (discrete) *stochastic process* is a real valued function

$$y : Z \times \Omega \to \mathbb{R}$$

such that for each fixed $t \in Z$, $y(t, \omega)$ is a random variable. The random variable corresponding to a fixed t is usually denoted by y_t in the following. The underlying probability space will usually not even be mentioned. In that case, it is understood that all the members y_t of a stochastic process are defined on the same probability space. Usually the stochastic process will also be denoted by y_t if the meaning of the symbol is clear from the context.

A stochastic process may be described by the joint distribution functions of all finite subcollections of y_t's, $t \in S \subset Z$. In practice, the complete system of distributions will often be unknown. Therefore, in the following chapters, we will often be concerned with the first and second moments of the distributions. In other words, we will be concerned with the means $E(y_t) = \mu_t$, the variances $E[(y_t - \mu_t)^2]$ and the covariances $E[(y_t - \mu_t)(y_s - \mu_s)]$.

A K-dimensional *vector stochastic process* or *multivariate stochastic process* is a function

$$y : Z \times \Omega \to \mathbb{R}^K,$$

where, for each fixed $t \in Z$, $y(t, \omega)$ is a K-dimensional random vector. Again we usually use the symbol y_t for the random vector corresponding to a fixed $t \in Z$. For simplicity, we also often denote the complete process by y_t. The particular meaning of the symbol should be clear from the context. With respect to the stochastic characteristics the same applies as for univariate processes. That is, the stochastic characteristics are summarized in the joint distribution functions of all finite subcollections of random vectors y_t. In practice, interest will often focus on the first and second moments of all random variables involved.

A realization of a (vector) stochastic process is a sequence (of vectors) $y_t(\omega)$, $t \in Z$, for a fixed ω. In other words, a realization of a stochastic process is a function $Z \to \mathbb{R}^K$ where $t \to y_t(\omega)$. A (multiple) time series is regarded as such a realization or possibly a finite part of such a realization, that is, it consists, for instance, of values (vectors) $y_1(\omega), \dots, y_T(\omega)$. The underlying stochastic process is said to have *generated* the (multiple) time series or it is called the *generating* or *generation process* of the time series or the *data generation process* (DGP). A time series $y_1(\omega), \dots, y_T(\omega)$ will usually be denoted by $y_1, \dots, y_T$ or simply by y_t just like the underlying stochastic process, if no confusion is possible. The number of observations, T, is called the *sample size* or *time series length.* With this terminology at hand, we may now return to the problem of specifying forecast functions.

1.3 Vector Autoregressive Processes

Because linear functions are relatively easy to deal with, it makes sense to begin with forecasts that are linear functions of past observations. Let us consider a univariate time series y_t and a forecast $h = 1$ period into the future. If $f(\cdot)$ in (1.1.1) is a linear function, we have

$$\widehat{y}_{T+1} = \nu + \alpha_1 y_T + \alpha_2 y_{T-1} + \cdots .$$

Assuming that only a finite number p, say, of past y values are used in the prediction formula, we get

$$\widehat{y}_{T+1} = \nu + \alpha_1 y_T + \alpha_2 y_{T-1} + \cdots + \alpha_p y_{T-p+1}. \tag{1.3.1}$$

Of course, the true value y_{T+1} will usually not be exactly equal to the forecast $\widehat{y}_{T+1}$. Let us denote the forecast error by $u_{T+1} := y_{T+1} - \widehat{y}_{T+1}$ so that

$$y_{T+1} = \widehat{y}_{T+1} + u_{T+1} = \nu + \alpha_1 y_T + \cdots + \alpha_p y_{T-p+1} + u_{T+1}. \tag{1.3.2}$$

Now, assuming that our numbers are realizations of random variables and that the same data generation law prevails in each period T, (1.3.2) has the form of an *autoregressive process*,

$$y_t = \nu + \alpha_1 y_{t-1} + \cdots + \alpha_p y_{t-p} + u_t, \tag{1.3.3}$$

where the quantities $y_t, y_{t-1}, \dots, y_{t-p}$, and u_t are now random variables. To actually get an autoregressive (AR) process we assume that the forecast errors u_t for different periods are uncorrelated, that is, u_t and u_s are uncorrelated for $s \neq t$. In other words, we assume that all useful information in the past y_t's is used in the forecasts so that there are no systematic forecast errors.

If a multiple time series is considered, an obvious extension of (1.3.1) would be

$$\begin{aligned} \widehat{y}_{k,T+1} &= \nu + \alpha_{k1,1} y_{1,T} + \alpha_{k2,1} y_{2,T} + \cdots + \alpha_{kK,1} y_{K,T} \\ &\quad + \cdots + \alpha_{k1,p} y_{1,T-p+1} + \cdots + \alpha_{kK,p} y_{K,T-p+1}, \\ &\qquad\qquad k = 1, \dots, K. \end{aligned} \tag{1.3.4}$$

To simplify the notation, let $y_t := (y_{1t}, \dots, y_{Kt})'$, $\widehat{y}_t := (\widehat{y}_{1t}, \dots, \widehat{y}_{Kt})'$, $\nu := (\nu_1, \dots, \nu_K)'$ and

$$A_i := \begin{bmatrix} \alpha_{11,i} & \dots & \alpha_{1K,i} \\ \vdots & \ddots & \vdots \\ \alpha_{K1,i} & \dots & \alpha_{KK,i} \end{bmatrix}.$$

Then (1.3.4) can be written compactly as

$$\widehat{y}_{T+1} = \nu + A_1 y_T + \dots + A_p y_{T-p+1}. \tag{1.3.5}$$

If the y_t's are regarded as random vectors, this predictor is just the optimal forecast obtained from a vector autoregressive model of the form

$$y_t = \nu + A_1 y_{t-1} + \dots + A_p y_{t-p} + u_t, \tag{1.3.6}$$

where the $u_t = (u_{1t}, \dots, u_{Kt})'$ form a sequence of independently identically distributed random K-vectors with zero mean vector.

Obviously such a model represents a tremendous simplification compared with the general form (1.1.2). Because of its simple structure, it enjoys great popularity in applied work. We will study this particular model in the following chapters in some detail.

1.4 Outline of the Following Chapters

In Part I of the book, consisting of the next four chapters, we will investigate some basic properties of stationary vector autoregressive (VAR) processes such as (1.3.6). Forecasts based on these processes are discussed and it is shown how VAR processes may be used for analyzing the dynamic structure of a system of variables. Throughout Chapter 2, it is assumed that the process under study is completely known including its coefficient matrices. In practice, for a given multiple time series, first a model of the DGP has to be specified and its parameters have to be estimated. Then the adequacy of the model is checked by various statistical tools and then the estimated model can be used for forecasting and dynamic or structural analysis. The main steps of a VAR analysis are presented in Figure 1.1 in a schematic way. Estimation and model specification are discussed in Chapters 3 and 4, respectively. In the former chapter the estimation of the VAR coefficients is considered and the consequences of using estimated rather than known processes for forecasting and economic analysis are explored. In Chapter 4, the specification and model checking stages of an analysis are considered. Criteria for determining the order p of a VAR process are given and possibilities for checking the assumptions underlying a VAR analysis are discussed.

In systems with many variables and/or large VAR order p, the number of coefficients is quite substantial. As a result the estimation precision will

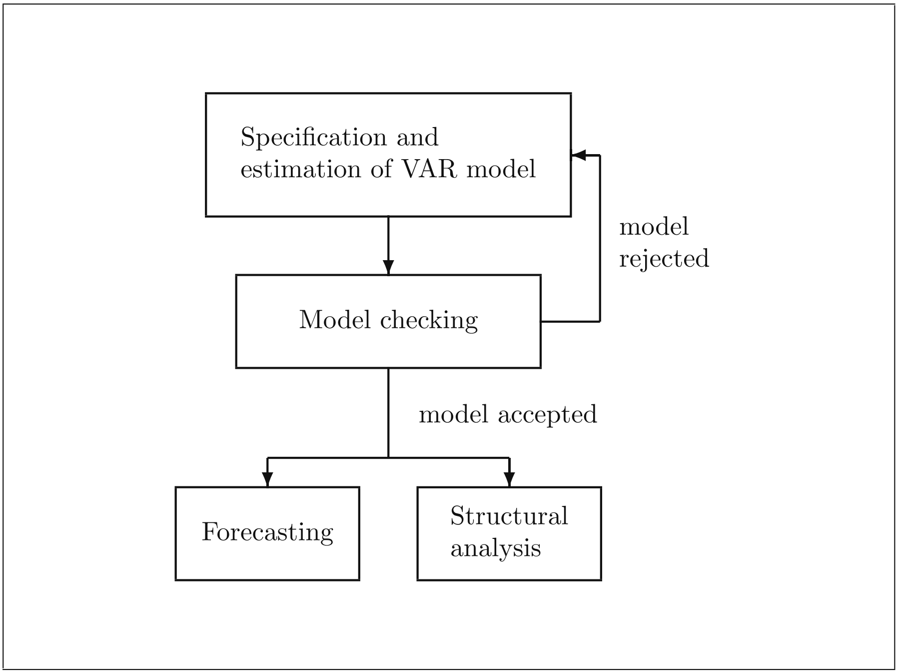

Fig. 1.1. VAR analysis.

be low if estimation is based on time series of the size typically available in economic applications. In order to improve the estimation precision, it is useful to place restrictions from nonsample sources on the parameters and thereby reduce the number of coefficients to be estimated. In Chapter 5, VAR processes with parameter constraints and restricted estimation are discussed. Zero restrictions, nonlinear constraints, and Bayesian estimation are treated.

In Part I, stationary processes are considered which have time invariant expected values, variances, and covariances. In other words, the first and second moments of the random variables do not change over time. In practice many time series have a trending behavior which is not compatible with such an assumption. This fact is recognized in Part II, where VAR processes with stochastic and deterministic trends are considered. Processes with stochastic trends are often called integrated and if two or more variables are driven by the same stochastic trend, they are called cointegrated. Cointegrated VAR processes have quite different properties from stationary ones and this has to be taken into account in the statistical analysis. The specific estimation, specification, and model checking procedures are discussed in Chapters 6–8.

The models discussed in Parts I and II are essentially reduced form models which capture the dynamic properties of the variables and are useful forecasting tools. For structural economic analysis, these models are often insufficient because different economic theories may be compatible with the same statistical reduced form model. In Chapter 9, it is discussed how to integrate structural information in stationary and cointegrated VAR models. In many econometric applications it is assumed that some of the variables are determined outside the system under consideration. In other words, they are exogenous or unmodelled variables. VAR processes with exogenous variables are considered in Chapter 10. In the econometrics literature such systems are often called systems of dynamic simultaneous equations. In the time series literature they are sometimes referred to as multivariate transfer function models. Together Chapters 9 and 10 constitute Part III of this volume.

In Part IV of the book, it is recognized that an upper bound p for the VAR order is often not known with certainty. In such a case, one may not want to impose any upper bound and allow for an infinite VAR order. There are two ways to make the estimation problem for the potentially infinite number of parameters tractable. First, it may be assumed that they depend on a finite set of parameters. This assumption leads to vector autoregressive moving average (VARMA) processes. Some properties of these processes, parameter estimation and model specification are discussed in Chapters 11–13 for the stationary case and in Chapter 14 for cointegrated systems. In the second approach for dealing with infinite order VAR processes, it is assumed that finite order VAR processes are fitted and that the VAR order goes to infinity with the sample size. This approach and its consequences for the estimators, forecasts, and structural analysis are discussed in Chapter 15 for both the stationary and the cointegrated cases.

In Part V, some special models and issues for multiple time series are studied. In Chapter 16, models for conditionally heteroskedastic series are considered and, in particular, multivariate generalized autoregressive conditionally heteroskedastic (MGARCH) processes are presented and analyzed. In Chapter 17, VAR processes with time varying coefficients are considered. The coefficient variability may be due to a one-time intervention from outside the system or it may result from seasonal variation. Finally, in Chapter 18, so-called state space models are introduced. The models represent a very general class which encompasses most of the models previously discussed and includes in addition VAR models with stochastically varying coefficients. A brief review of these and other important models for multiple time series is given. The Kalman filter is presented as an important tool for dealing with state space models.

The reader is assumed to be familiar with vectors and matrices. The rules used in the text are summarized in Appendix A. Some results on the multivariate normal and related distributions are listed in Appendix B and stochastic convergence and some asymptotic distribution theory are reviewed in Appendix C. In Appendix D, a brief outline is given of the use of simulation techniques in evaluating properties of estimators and test statistics. Although it is not necessary for the reader to be familiar with all the particular rules and propositions listed in the appendices, it is implicitly assumed in the following chapters that the reader has knowledge of the basic terms and results.