18

State Space Models

18.1 Background

State space models may be regarded as generalizations of the models considered so far. They have been used extensively in system theory, the physical sciences, and engineering. The terminology is therefore largely from these fields. The general idea behind these models is that an observed (multiple) time series $y_1, \ldots, y_T$ depends upon a possibly unobserved state z_t which is driven by a stochastic process. The relation between y_t and z_t is described by the *observation* or *measurement equation*

$$y_t = \mathbf{H}_t z_t + v_t, \tag{18.1.1}$$

where $\mathbf{H}_t$ is a matrix that may also depend on the period of time, t, and v_t is the *observation error* which is typically assumed to be a noise process. The *state vector* or state of nature is generated as

$$z_t = \mathbf{B}_{t-1} z_{t-1} + w_{t-1} \tag{18.1.2}$$

which is often called the *transition equation* because it describes the transition of the state of nature from period $t-1$ to period t. The matrix $\mathbf{B}_t$ is a coefficient matrix that may depend on t and w_t is an error process. The system (18.1.1)/(18.1.2) is one form of a *state space model*.

The following example from Meinhold & Singpurwalla (1983) may illustrate the related concepts. Suppose we wish to trace a satellite's orbit. The state vector z_t may then consist of the position and the speed of the satellite in period t with respect to the center of the earth. The state cannot be measured directly but, for example, the distance from a certain observatory may be measured. These measurements constitute the observed vectors y_t. As another example, consider the income of an individual which may depend on unobserved factors such as intelligence, special abilities, special interests and so on. In this case, the state vector consists of the variables that describe the abilities of the person and y_t is his or her observed income.

H. Lütkepohl, *New Introduction to Multiple Time Series Analysis*

The reader may recall that all the models considered so far have been written in a form similar to (18.1.1)/(18.1.2) at some stage. For instance, our standard (zero mean) VAR(p) model can be written in VAR(1) form as

$$Y_t = \mathbf{A}Y_{t-1} + U_t, \tag{18.1.3}$$

where

$$Y_t := \begin{bmatrix} y_t \\ \vdots \\ y_{t-p+1} \end{bmatrix}, \quad U_t := \begin{bmatrix} u_t \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \quad \text{etc..}$$

Defining $z_t := Y_t$, $\mathbf{B}_t := \mathbf{A}$, and $w_{t-1} := U_t$, Equation (18.1.3) may be viewed as the transition equation of a state space model. The corresponding measurement equation is

$$y_t = [I_K : 0 : \cdots : 0]Y_t$$

with $\mathbf{H}_t := [I_K : 0 : \cdots : 0]$ and $v_t := 0$.

In the next section, we will introduce a slightly more general version of a state space model, we will review many of the previous models, and we will cast them into state space form. As we have seen, the representations of the models used in the previous chapters are useful for many purposes. There are occasions, however, where a state space representation makes life easier. We have actually used state space representations of some models without explicitly mentioning this fact. We will also consider some further models that have been discussed in the literature and which may be set up as special cases of state space models. Thereby we will give an overview of a number of important models that have been considered in the multiple time series literature.

In Section 18.3, we will discuss the Kalman filter which is an extremely useful tool in the analysis of state space models. Given the observable vectors y_t, it provides estimates of the state vectors and measures of the precision of these estimates. In a situation where the state vector consists of unobservable variables, such estimates may be of interest. In a system such as (18.1.1)/(18.1.2), the matrices $\mathbf{B}_t$ and $\mathbf{H}_t$ and the covariance matrices of v_t and w_t will often depend on unknown parameters. The Kalman filter is also helpful in estimating these parameters. This issue will be discussed in Section 18.4.

In this chapter, we will just give a brief introduction to some basic concepts related to state space models and the Kalman filter. Various textbooks exist that provide broader introductions to the topic and a more in-depth discussion. Examples are Jazwinski (1970), Anderson & Moore (1979), Hannan & Deistler (1988), Aoki (1987), and Harvey (1989).

18.2 State Space Models

18.2.1 The Model Setup

As mentioned in the previous section, a state space model consists of a *transition* or *system equation*

$$z_{t+1} = \mathbf{B}_t z_t + \mathbf{F}_t x_t + w_t, \quad t = 0, 1, 2, \ldots,$$

or, equivalently,

$$z_t = \mathbf{B}_{t-1} z_{t-1} + \mathbf{F}_{t-1} x_{t-1} + w_{t-1}, \quad t = 1, 2, \ldots, \tag{18.2.1}$$

and a *measurement* or *observation equation*

$$y_t = \mathbf{H}_t z_t + \mathbf{G}_t x_t + v_t, \quad t = 1, 2, \ldots. \tag{18.2.2}$$

Here

y_t is a $(K \times 1)$ vector of observable *output* or *endogenous variables*,
z_t is an $(N \times 1)$ *state vector* or the *state of nature*,
x_t is an $(M \times 1)$ vector of observable *inputs* or *instruments* or *policy variables*,
v_t is a $(K \times 1)$ vector of *observation* or *measurement errors* or *noise*,
w_t is an $(N \times 1)$ vector of *system* or *transition equation errors* or *noise*,
$\mathbf{H}_t$ is a $(K \times N)$ *measurement matrix*,
$\mathbf{G}_t$ is a $(K \times M)$ *input matrix* of the observation equation,
$\mathbf{B}_t$ is an $(N \times N)$ *transition* or *system matrix*,

and

$\mathbf{F}_t$ is a $(N \times M)$ *input matrix* of the transition equation.

The matrices $\mathbf{H}_t$, $\mathbf{G}_t$, $\mathbf{B}_t$, and $\mathbf{F}_t$ are assumed to be known at time t. Although they are in general allowed to vary, at least some of them will often be time invariant. In practice, at least some of the elements of these matrices are usually unknown and have to be estimated. This issue is deferred to Section 18.4. It is perhaps noteworthy that the process generating the z_t's and, hence, also the y_t's is assumed to be started from an *initial state* z_0 and a given *initial input* x_0.

To complete the description of the model, we make the following stochastic assumptions for the noise processes and the initial state:

The joint process

$$\begin{bmatrix} w_t \\ v_t \end{bmatrix}$$

is a zero mean, serially uncorrelated noise process with possibly time varying covariance matrices

$$\begin{bmatrix} \Sigma_{w_t} & \Sigma_{w_t v_t} \\ \Sigma_{v_t w_t} & \Sigma_{v_t} \end{bmatrix}.$$

The initial state z_0 is uncorrelated with w_t, v_t for all t and has a distribution with mean μ_{z_0} and covariance matrix Σ_{z_0}. The input sequence $x_0, x_1, \ldots$ is assumed to be nonstochastic for simplicity. If the observed inputs are actually stochastic, the analysis is assumed to be conditional on a given sequence of inputs.

With these assumptions we can derive stochastic properties of the states and the system outputs. Successive substitution in (18.2.1) implies

$$z_t = \mathbf{\Phi}_{t,t} z_0 + \sum_{i=1}^{t} \mathbf{\Phi}_{i-1,t}(\mathbf{F}_{t-i} x_{t-i} + w_{t-i}), \tag{18.2.3}$$

where

$$\mathbf{\Phi}_{0,t} := I_N \quad \text{and} \quad \mathbf{\Phi}_{i,t} := \prod_{j=1}^{i} \mathbf{B}_{t-j}, \quad i = 1, 2, \ldots$$

(see also Section 17.2.1). Hence,

$$\mu_{z_t} := E(z_t) = \mathbf{\Phi}_{t,t}\mu_{z_0} + \sum_{i=1}^{t} \mathbf{\Phi}_{i-1,t}\mathbf{F}_{t-i} x_{t-i} \tag{18.2.4}$$

and

$$\begin{aligned} \mathrm{Cov}(z_t, z_{t+h}) &= E[(z_t - \mu_{z_t})(z_{t+h} - \mu_{z_{t+h}})'] \\ &= \mathbf{\Phi}_{t,t}\Sigma_{z_0}\mathbf{\Phi}'_{t+h,t+h} + \sum_{i=1}^{t} \mathbf{\Phi}_{i-1,t}\Sigma_{w_{t-i}}\mathbf{\Phi}'_{h+i-1,t+h}. \end{aligned} \tag{18.2.5}$$

Under the aforementioned stochastic assumptions, it is also easy to derive the means and covariance matrices of the output process:

$$\mu_{y_t} := E(y_t) = \mathbf{H}_t E(z_t) + \mathbf{G}_t x_t$$

and

$$\mathrm{Cov}(y_t, y_{t+h}) = \mathbf{H}_t \mathrm{Cov}(z_t, z_{t+h})\mathbf{H}'_t \quad \text{for } h \neq 0.$$

Generally, the means and autocovariances of the y_t's are obviously not time invariant. Thus, in general, y_t is a nonstationary process.

We will now consider various special cases of state space models which are obtained by specific definitions of the state vector, the inputs, the noise processes, and the matrices $\mathbf{H}_t$, $\mathbf{G}_t$, $\mathbf{B}_t$, and $\mathbf{F}_t$. These matrices and the noise covariance matrices will often not depend on t, in which case we will suppress the subscript for notational simplicity.

A Finite Order VAR Process

Although we have mentioned earlier how to cast a VAR(p) process,

$$y_t = \nu + A_1 y_{t-1} + \cdots + A_p y_{t-p} + u_t, \tag{18.2.6}$$

in state space form, it may be useful to consider this model again because it illustrates that often different state space models can represent a particular process. One possible state space representation is obtained by defining

$$Y_t := \begin{bmatrix} y_t \\ \vdots \\ y_{t-p+1} \end{bmatrix}, \quad \boldsymbol{\nu} := \begin{bmatrix} \nu \\ 0 \\ \vdots \\ 0 \end{bmatrix},$$

$$\mathbf{A} := \begin{bmatrix} A_1 & \dots & A_{p-1} & A_p \\ I_K & & 0 & 0 \\ & \ddots & \vdots & \vdots \\ 0 & \dots & I_K & 0 \end{bmatrix}, \quad U_t := \begin{bmatrix} u_t \\ 0 \\ \vdots \\ 0 \end{bmatrix}. \tag{18.2.7}$$

Hence,

$$Y_t = \mathbf{A} Y_{t-1} + \boldsymbol{\nu} + U_t, \tag{18.2.8}$$

$$y_t = [I_K : 0 : \cdots : 0] Y_t \tag{18.2.9}$$

is a state space model with state vector $z_t := Y_t$, $\mathbf{B} := \mathbf{A}$, $\mathbf{F} := \boldsymbol{\nu}$, $x_t := 1$, $w_t := U_{t+1}$, $\mathbf{H} := [I_K : 0 : \cdots : 0]$, $\mathbf{G} := 0$, $v_t := 0$.

An alternative possibility is to define the state vector as

$$z_t := \begin{bmatrix} 1 \\ y_t \\ \vdots \\ y_{t-p+1} \end{bmatrix}$$

and choose

$$\mathbf{B} := \begin{bmatrix} 1 & 0 & \dots & 0 & 0 \\ \nu & A_1 & \dots & A_{p-1} & A_p \\ 0 & I_K & & 0 & 0 \\ \vdots & & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & I_K & 0 \end{bmatrix} \quad \text{and} \quad w_t := \begin{bmatrix} 0 \\ u_{t+1} \\ 0 \\ \vdots \\ 0 \end{bmatrix},$$

so that

$$z_{t+1} = \mathbf{B} z_t + w_t$$

and

$$y_t = [0 : I_K : 0 : \cdots : 0]z_t,$$

which describes the same process as (18.2.8)/(18.2.9). It may be worth pointing out that in the present framework, the process is assumed to be started at time $t = 1$ with initial values $z_0 = [1, y_0', \ldots, y_{-p+1}']'$, while we have assumed an infinite past of the process in some previous chapters.

A VARMA(p, q) Process

One state space representation of the VARMA(p, q) process

$$y_t = \nu + A_1 y_{t-1} + \cdots + A_p y_{t-p} + u_t + M_1 u_{t-1} + \cdots + M_q u_{t-q} \quad (18.2.10)$$

is known from Chapter 11, Section 11.3.2. It is obtained by choosing a state vector

$$z_t := \begin{bmatrix} y_t \\ \vdots \\ y_{t-p+1} \\ u_t \\ \vdots \\ u_{t-q+1} \end{bmatrix}, \quad \text{transition noise } w_t := \begin{bmatrix} u_{t+1} \\ 0 \\ \vdots \\ 0 \\ u_{t+1} \\ 0 \\ \vdots \\ 0 \end{bmatrix},$$

an input sequence $x_t := 1$ as in (18.2.8), $\mathbf{B} := \mathbf{A}$ from Chapter 11, Equation (11.3.8), $\mathbf{F} := \boldsymbol{\nu}$ defined similarly as in (18.2.8), $\mathbf{H} := [I_K : 0 : \cdots : 0]$, $\mathbf{G} := 0$, and $v_t := 0$. For many purposes, this form is not the most useful state space representation of a VARMA model. Other state space representations are given by Aoki (1987), Hannan & Deistler (1988), and Wei (1990).

The VARX Model

The VARX model

$$y_t = A_1 y_{t-1} + \cdots + A_p y_{t-p} + B_0 x_t + \cdots + B_s x_{t-s} + u_t \quad (18.2.11)$$

considered in Chapter 10 is easily cast in state space form by choosing the state vector

$$z_t := \begin{bmatrix} y_t \\ \vdots \\ y_{t-p+1} \\ x_t \\ \vdots \\ x_{t-s+1} \end{bmatrix},$$

and the transition equation

$$z_t = \left[\begin{array}{cccc|ccccc} A_1 & \dots & A_{p-1} & A_p & B_1 & \dots & B_{s-1} & B_s \\ I & & 0 & 0 & 0 & \dots & 0 & 0 \\ & \ddots & \vdots & \vdots & \vdots & & \vdots & \vdots \\ 0 & \dots & I & 0 & 0 & \dots & 0 & 0 \\ \hline & & & & 0 & \dots & 0 & 0 \\ & & 0 & & I & & 0 & 0 \\ & & & & & \ddots & \vdots & \vdots \\ & & & & 0 & \dots & I & 0 \end{array}\right] z_{t-1}$$

$$+ \left[\begin{array}{c} B_0 \\ 0 \\ \vdots \\ 0 \\ I \\ 0 \\ \vdots \\ 0 \end{array}\right] x_t + \left[\begin{array}{c} u_t \\ 0 \\ \vdots \\ 0 \end{array}\right]. \tag{18.2.12}$$

The corresponding observation equation is

$$y_t = [I_K : 0 : \cdots : 0] z_t.$$

It is also possible to extend the model so as to allow for a finite order MA(q) error process in (18.2.11) (see Problem 18.1).

Systematic Sampling and Aggregation

Suppose that annual data is available whereas a decision maker is interested in, say, quarterly figures. Let η_{it} be an $(M \times 1)$ vector of variables associated with the i-th quarter of year t and suppose the vector of all quarterly variables associated with year t,

$$\eta_t := \left[\begin{array}{c} \eta_{1t} \\ \eta_{2t} \\ \eta_{3t} \\ \eta_{4t} \end{array}\right],$$

is generated by the VAR(p) process

$$\eta_t = A_1 \eta_{t-1} + \cdots + A_p \eta_{t-p} + u_t.$$

Then we may define a state vector

$$z_t := \begin{bmatrix} \eta_t \\ \vdots \\ \eta_{t-p+1} \end{bmatrix}$$

and a transition equation

$$z_t = \mathbf{A} z_{t-1} + U_t,$$

where $\mathbf{A}$ and U_t are the same quantities as in (18.2.7). If the yearly values are obtained by adding (aggregating) the quarterly figures, the observation equation is

$$y_t = [I_M : I_M : I_M : I_M : 0 : \cdots : 0] z_t,$$

where M is the dimension of η_{it}. Alternatively, if the annual figures are obtained by systematic sampling, that is, by taking, say, the fourth quarter values as the annual figures, the observation equation is

$$y_t = [0 : 0 : 0 : I_M : 0 : \cdots : 0] z_t.$$

Extensions of this framework to the case where η_t is generated by a VARMA or VARX process are straightforward. For applications of state space models in aggregation and systematic sampling problems see Nijman (1985), Harvey (1984), Harvey & Pierse (1984), Jones (1980), Ansley & Kohn (1983).

The examples considered so far have in common that the system matrices $\mathbf{H}$, $\mathbf{G}$, $\mathbf{B}$, and $\mathbf{F}$ are all time invariant and the state vector consists of at least some observed or observable variables. In contrast, the state vector is unobservable in the next two examples while the system matrices remain time invariant.

Structural Time Series Models

In a structural time series model, the observed time series is viewed as a sum of unobserved components such as a trend, a seasonal component, and an irregular component (see, e.g., Kitagawa (1981), Harvey & Todd (1983), Harvey (1989)). For instance, for a univariate time series $y_1, \ldots, y_T$, the structural model may have the form

$$y_t = \mu_t + \gamma_t + u_t, \tag{18.2.13}$$

where μ_t is a trend component and γ_t is a seasonal component. Harvey & Todd (1983) assume a local approximation to a linear trend function for which both the level and the slope are shifting. They postulate a process

$$\mu_t = \mu_{t-1} + \beta_{t-1} + \eta_t \quad \text{with } \beta_t = \beta_{t-1} + \xi_t \tag{18.2.14}$$

as the trend generation mechanism. Here η_t and ξ_t are assumed to be white noise processes. This trend model is a mixture of two random walks which are

discussed in Chapter 6, Section 6.1. For the seasonal component, it is assumed that the sum over the seasonal factors of a full year is approximately zero,

$$\gamma_t = -\sum_{j=1}^{s-1} \gamma_{t-j} + \omega_t, \tag{18.2.15}$$

where s is the number of seasons and ω_t is white noise. The three white noise processes η_t, ξ_t, and ω_t are assumed to be independent.

This model can be set up in state space form by defining the state vector to be

$$z_t := \begin{bmatrix} \mu_t \\ \beta_t \\ \gamma_t \\ \vdots \\ \gamma_{t-s+2} \end{bmatrix}$$

and, hence, the transition equation is

$$z_t = \left[\begin{array}{cc|ccccc} 1 & 1 & & & & & \\ 0 & 1 & & & 0 & & \\ \hline & & & -1 & \dots & -1 & -1 \\ & & & 1 & & 0 & 0 \\ & 0 & & & \ddots & \vdots & \vdots \\ & & & 0 & \dots & 1 & 0 \end{array}\right] z_{t-1} + \begin{bmatrix} \eta_t \\ \xi_t \\ \omega_t \\ 0 \\ \vdots \\ 0 \end{bmatrix}. \tag{18.2.16}$$

The corresponding measurement equation is

$$y_t = [1, 0, 1, 0, \dots, 0] z_t + u_t. \tag{18.2.17}$$

It may be worth noting that these models can be seen to describe special integrated ARMA processes. Hence, it is, of course, not surprising that they can be cast in state space form. Multivariate generalizations of this model are possible (see Harvey (1987) and Proietti (2002)).

Factor Analytic Models

In a classical factor analytic setting, it is assumed that a set of K observed variables y_t depends linearly on $N < K$ unobserved common factors f_t and on individual or idiosyncratic components u_t. In other words,

$$y_t = \mathbf{L} f_t + u_t, \tag{18.2.18}$$

where $\mathbf{L}$ is a $(K \times N)$ matrix of *factor loadings* and the components of u_t are typically assumed to be uncorrelated, that is, Σ_u is a diagonal matrix (Anderson (1984), Morrison (1976)). One objective of a factor analysis is the

construction or estimation of the unobserved factors f_t. We may view (18.2.18) as the measurement equation of a state space model and, if the factors f_t and f_s are independent for $t \neq s$, we may specify a trivial transition equation $f_t = w_{t-1}$.

However, if y_t consists of time series variables, it may be more reasonable to assume that the factors are autocorrelated. For example, they may be generated by a VAR or VARMA process. Also the idiosyncratic components u_t may be autocorrelated. *Dynamic factor analytic models* of this type were considered, for instance, by Sargent & Sims (1977), Geweke (1977), and Engle & Watson (1981). Assuming that

$$f_t = A_1 f_{t-1} + \cdots + A_p f_{t-p} + \eta_t$$

and

$$u_t = C_1 u_{t-1} + \cdots + C_q u_{t-q} + \varepsilon_t,$$

where η_t and ε_t are white noise processes, a state space model can be set up by specifying a state vector,

$$z_t := \begin{bmatrix} f_t \\ \vdots \\ f_{t-p+1} \\ u_t \\ \vdots \\ u_{t-q+1} \end{bmatrix},$$

and a transition equation

$$z_t = \left[\begin{array}{cccc|cccc} A_1 & \dots & A_{p-1} & A_p & & & & \\ I & & 0 & 0 & & & 0 & \\ & \ddots & \vdots & \vdots & & & & \\ 0 & \dots & I & 0 & & & & \\ \hline & & & & C_1 & \dots & C_{q-1} & C_q \\ & & 0 & & I & & 0 & 0 \\ & & & & & \ddots & \vdots & \vdots \\ & & & & 0 & \dots & I & 0 \end{array}\right] z_{t-1} + \begin{bmatrix} \eta_t \\ 0 \\ \vdots \\ 0 \\ \varepsilon_t \\ 0 \\ \vdots \\ 0 \end{bmatrix}. \tag{18.2.19}$$

The corresponding measurement equation is

$$y_t = \mathbf{L} f_t + u_t = [\mathbf{L} : 0 : \cdots : 0 : I_K : 0 : \cdots : 0] z_t. \tag{18.2.20}$$

An extension to the case where f_t and u_t are generated by VARMA processes is left to the reader (see Problem 18.2). If exogenous variables are added

to the original model (18.2.18) and, in addition, the factors are dynamic processes, we obtain the dynamic MIMIC models of Engle & Watson (1981). More recent references on dynamic factor models include Stock & Watson (2002a, b) and Forni, Hallin, Lippi & Reichlin (2000).

In all the previous examples the system matrices $\mathbf{H}_t$, $\mathbf{G}_t$, $\mathbf{B}_t$, and $\mathbf{F}_t$ are time invariant. We will now consider models where at least some elements of these matrices vary through time.

VARX Models with Systematically Varying Coefficients

We extend the varying coefficients VAR models of Chapter 17 slightly by adding further "exogenous" variables and assuming that a given multiple time series is generated according to

$$y_t = A_{1,t}y_{t-1} + \cdots + A_{p,t}y_{t-p} + F_t x_t + u_t. \tag{18.2.21}$$

The vector x_t may simply include an intercept term or seasonal dummies. It may also include other deterministic terms and even lags of exogenous variables. Because we are assuming that the input variables of the state space model are nonstochastic, we restrict x_t to be a deterministic sequence, however. Using

$$Y_t := \begin{bmatrix} y_t \\ \vdots \\ y_{t-p+1} \end{bmatrix}, \quad \mathbf{A}_t := \begin{bmatrix} A_{1,t} & \dots & A_{p-1,t} & A_{p,t} \\ I & & 0 & 0 \\ & \ddots & \vdots & \vdots \\ 0 & \dots & I & 0 \end{bmatrix},$$

$$\mathbf{F}_t := \begin{bmatrix} F_t \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \quad \text{and} \quad w_{t-1} := \begin{bmatrix} u_t \\ 0 \\ \vdots \\ 0 \end{bmatrix},$$

gives a transition equation

$$Y_t = \mathbf{A}_t Y_{t-1} + \mathbf{F}_t x_t + w_{t-1} \tag{18.2.22}$$

and a measurement equation

$$y_t = [I_K : 0 : \cdots : 0]Y_t. \tag{18.2.23}$$

Obviously, the transition matrix $\mathbf{B}_{t-1} := \mathbf{A}_t$ and the input matrix $\mathbf{F}_t$ of the transition equation may be time varying in this state space model.

Random Coefficients VARX Models

So far all the original models either have time invariant, constant coefficients or, as in the previous example, systematically varying coefficients. We will now consider models with random coefficients and demonstrate how they can be cast in state space form. Let us begin with a simple multivariate regression model of the form

$$y_t = C_t x_t + v_t = (x_t' \otimes I)\text{vec}(C_t) + v_t. \tag{18.2.24}$$

Assuming that the parameter vector $\boldsymbol{\gamma}_t := \text{vec}(C_t)$ is generated by a VAR(q) process,

$$\boldsymbol{\gamma}_t = \nu + B_1\boldsymbol{\gamma}_{t-1} + \cdots + B_q\boldsymbol{\gamma}_{t-q} + u_t, \tag{18.2.25}$$

we may define the state vector as

$$z_t := \begin{bmatrix} \boldsymbol{\gamma}_t \\ \vdots \\ \boldsymbol{\gamma}_{t-q+1} \end{bmatrix}$$

and get a state space model with the following transition and measurement equations, respectively:

$$z_t = \begin{bmatrix} B_1 & \dots & B_{q-1} & B_q \\ I & & 0 & 0 \\ & \ddots & \vdots & \vdots \\ 0 & \dots & I & 0 \end{bmatrix} z_{t-1} + \begin{bmatrix} \nu \\ 0 \\ \vdots \\ 0 \end{bmatrix} + \begin{bmatrix} u_t \\ 0 \\ \vdots \\ 0 \end{bmatrix},$$

$$y_t = [x_t' \otimes I : 0 : \cdots : 0]z_t + v_t.$$

Obviously, the measurement matrix,

$$\mathbf{H}_t := [x_t' \otimes I : 0 : \cdots : 0],$$

may be time varying. It may, in fact, be random if the x_t are stochastic variables. Such an assumption is mandatory if x_t contains lagged y_t variables. To see this point more clearly, let us explicitly introduce lagged y_t's in (18.2.24):

$$\begin{aligned} y_t &= \mathbf{A}_t Y_{t-1} + C_t x_t + v_t \\ &= (Y_{t-1}' \otimes I)\text{vec}(\mathbf{A}_t) + (x_t' \otimes I)\text{vec}(C_t) + v_t, \end{aligned} \tag{18.2.26}$$

where $\mathbf{A}_t := [A_{1t}, \dots, A_{pt}]$ and $Y_{t-1}' := [y_{t-1}', \dots, y_{t-p}']$. Now suppose that $\boldsymbol{\gamma}_t := \text{vec}(C_t)$ is generated by the VAR(q) process (18.2.25) and $\boldsymbol{\alpha}_t = \text{vec}(\mathbf{A}_t)$ is driven by a VAR(r) process

$$\boldsymbol{\alpha}_t = D_1\boldsymbol{\alpha}_{t-1} + \cdots + D_r\boldsymbol{\alpha}_{t-r} + \eta_t, \tag{18.2.27}$$

which is assumed to be independent of $\boldsymbol{\gamma}_t$. Defining the state vector as

$$z_t := \begin{bmatrix} \boldsymbol{\gamma}_t \\ \vdots \\ \boldsymbol{\gamma}_{t-q+1} \\ \boldsymbol{\alpha}_t \\ \vdots \\ \boldsymbol{\alpha}_{t-r+1} \end{bmatrix},$$

the following state space model is obtained:

$$z_t = \left[\begin{array}{cccc|cccc} B_1 & \dots & B_{q-1} & B_q & & & & \\ I & & 0 & 0 & & & & \\ & \ddots & \vdots & \vdots & & & 0 & \\ 0 & \dots & I & 0 & & & & \\ \hline & & & & D_1 & \dots & D_{r-1} & D_r \\ & & 0 & & I & & 0 & 0 \\ & & & & & \ddots & \vdots & \vdots \\ & & & & 0 & \dots & I & 0 \end{array}\right] z_{t-1}$$

$$+ \begin{bmatrix} \nu \\ 0 \\ \vdots \\ 0 \\ \vdots \\ 0 \end{bmatrix} + \begin{bmatrix} u_t \\ 0 \\ \vdots \\ 0 \\ \eta_t \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \tag{18.2.28}$$

$$y_t = [x_t' \otimes I : 0 : \cdots : 0 : Y_{t-1}' \otimes I : 0 : \cdots : 0] z_t + v_t. \tag{18.2.29}$$

Further extensions of this model are possible. For instance, $\boldsymbol{\alpha}_t$ and $\boldsymbol{\gamma}_t$ may be individually or jointly generated by a VARMA rather than a finite order VAR process. Moreover, input variables with constant coefficients could appear in (18.2.26). These extensions are left to the reader (see Problem 18.3).

The number of publications on random coefficients models is vast in both the econometrics and the time series literature. Famous examples from the earlier econometrics literature on the topic are Hildreth & Houck (1968), Swamy (1971), and Cooley & Prescott (1973, 1976). Surveys of the earlier literature were given by Chow (1984) and Nicholls & Pagan (1985). Both of these articles include extensive reference lists. For a more recent overview see also Swamy & Tavlas (2001). On the time series side, a number of references can be found in the monograph by Nicholls & Quinn (1982). Other important work on the topic includes the article by Doan et al. (1984) who investigated the potential of random coefficients VAR models with Bayesian restrictions for econometric time series analysis.

18.2.2 More General State Space Models

There are also time series models that do not fall into the state space framework considered so far. Therefore it may be worth pointing out that more general nonlinear state space models have been studied in recent publications. A very general setup has the form

$$z_{t+1} = \mathbf{b}_t(z_t, x_t, w_t, \boldsymbol{\delta}_1) \tag{18.2.30}$$

for the transition equation and

$$y_t = \mathbf{h}_t(z_t, x_t, v_t, \boldsymbol{\delta}_2) \tag{18.2.31}$$

for the measurement equation. In other words, the functional dependence between the inputs, the states, and the output variables may be of a general nonlinear form and also the transition from one state to the next is described by a more general function than previously. Here $\boldsymbol{\delta}_1$ and $\boldsymbol{\delta}_2$ are vectors of parameters.

Bilinear time series models are examples for which the linear state space framework is too narrow. A very simple univariate bilinear time series model has the form

$$y_t = \alpha y_{t-1} + u_t + \beta y_{t-1} u_{t-1},$$

where u_t is univariate white noise. The product term $\beta y_{t-1} u_{t-1}$ distinguishes this model from a linear specification. Bilinear models have been found useful in modelling nonnormal phenomena (see, e.g., Granger & Andersen (1978)).

A more general multivariate bilinear time series model may be specified as follows:

$$\begin{aligned} y_t &= A_1 y_{t-1} + \cdots + A_p y_{t-p} + u_t + M_1 u_{t-1} + \cdots + M_q u_{t-q} \\ &\quad + \sum_{i=1}^{r} \sum_{j=1}^{s} C_{ij} \operatorname{vec}(y_{t-i} u'_{t-j}). \end{aligned} \tag{18.2.32}$$

Assuming, without loss of generality, that $p \geq r$ and $q \geq s$ and defining

$$z_t := \begin{bmatrix} y_t \\ \vdots \\ y_{t-p+1} \\ u_t \\ \vdots \\ u_{t-q+1} \end{bmatrix},$$

$$\mathbf{B} := \left[\begin{array}{cccc|ccc} A_1 & \dots & A_{p-1} & A_p & M_1 & \dots & M_{q-1} & M_q \\ I & & 0 & 0 & 0 & \dots & 0 & 0 \\ & \ddots & \vdots & \vdots & \vdots & & \vdots & \vdots \\ 0 & \dots & I & 0 & 0 & \dots & 0 & 0 \\ \hline & & & & 0 & \dots & 0 & 0 \\ & & & & I & & 0 & 0 \\ & & 0 & & & \ddots & \vdots & \vdots \\ & & & & 0 & \dots & I & 0 \end{array}\right],$$

$$w_t := \begin{bmatrix} u_{t+1} \\ 0 \\ \vdots \\ 0 \\ u_{t+1} \\ 0 \\ \vdots \\ 0 \end{bmatrix},$$

and a matrix $\mathbf{C}$ which contains the elements of the C_{ij} matrices in a suitable arrangement, we get a bilinear state space model of the form

$$z_{t+1} = \mathbf{B}z_t + w_t + \mathbf{C}\text{vec}(z_t z_t'), \tag{18.2.33}$$

$$y_t = [I_K : 0 : \cdots : 0]z_t. \tag{18.2.34}$$

Obviously, the transition equation involves a nonlinear term, namely $\text{vec}(z_t z_t')$. Hence, (18.2.33)/(18.2.34) is an example of a nonlinear state space system.

The work of Granger & Andersen (1978) and others on univariate bilinear models has stimulated investigations in this area. Much of the earlier work is documented in a monograph by Subba Rao & Gabr (1984). More recent work on multivariate bilinear models includes Stensholt & Tjøstheim (1987) and Liu (1989).

With all these examples we have not nearly exhausted the range of models that have been used and studied in the recent time series literature. Important omissions are threshold autoregressive models analyzed by Tong (1983) and exponential autoregressive models introduced by Ozaki (1980) and Haggan & Ozaki (1980). A general nonlinear model class was considered by Priestley (1980) and reviews of many nonlinear models and extensive lists of references were given by Priestley (1988), Anděl (1989), and Granger & Teräsvirta (1993).

18.3 The Kalman Filter

The Kalman filter was originally developed by Kalman (1960) and Kalman & Bucy (1961). It is a tool to recursively estimate the states z_t, given observations $y_1, \dots, y_T$ of the output variables. Under normality assumptions, the

estimator of the state produced by the Kalman filter is the conditional expectation $E(z_t|y_1, \ldots, y_t)$. The Kalman filter also provides the conditional covariance matrix $\mathrm{Cov}(z_t|y_1, \ldots, y_t)$ which may serve as a measure for estimation or prediction uncertainty. Of course, for $t > T$, the estimator $E(z_t|y_1, \ldots, y_T)$ is a forecast or prediction at origin T, in the terminology of the previous chapters. The computation of the estimators $E(z_t|y_1, \ldots, y_t)$, $t = 1, \ldots, T$, is called *filtering* to distinguish it from the forecasting problem.

In some of the examples of Section 18.2, estimation of the state vectors is of obvious interest, for instance, if the state vector consists of time varying coefficients or if the state vector contains the unobserved factors of a dynamic factor analytic model. In other cases, where the state vector is not of foremost interest or where it consists of observable variables, the conditional means and covariance matrices can still be useful in evaluating the likelihood function, for example. We will return to this point in Section 18.4. Now the Kalman filter recursions will be presented.

18.3.1 The Kalman Filter Recursions

Assumptions for the State Space Model

We assume a state space model with transition equation

$$z_t = \mathbf{B}z_{t-1} + \mathbf{F}x_{t-1} + w_{t-1} \tag{18.3.1}$$

and with measurement equation

$$y_t = \mathbf{H}_t z_t + \mathbf{G}x_t + v_t \tag{18.3.2}$$

for $t = 1, 2, \ldots$. Note that both input matrices and the transition matrix are assumed to be time invariant and known. This condition is satisfied in most of the example models of Section 18.2. The measurement matrices $\mathbf{H}_t$ are assumed to be known and nonstochastic at time t. This assumption does not exclude lagged output variables from $\mathbf{H}_t$ because the past output variables are given at time t. The input sequence x_t, $t = 0, 1, \ldots$, is again assumed to be nonstochastic for simplicity. The noise processes w_t and v_t are independent. They are both Gaussian with time invariant covariances,

$$\begin{aligned} w_t &\sim \mathcal{N}(0, \Sigma_w), \qquad t = 0, 1, \ldots, \\ v_t &\sim \mathcal{N}(0, \Sigma_v), \qquad t = 1, 2, \ldots. \end{aligned}$$

Also the initial state is Gaussian, $z_0 \sim \mathcal{N}(\mu_0, \Sigma_0)$, and it is assumed to be independent of v_t, w_{t-1}, $t = 1, \ldots$. The initial state may be a constant, nonstochastic vector in which case $\Sigma_0 = 0$.

With the exception of the normality assumption, the foregoing conditions are satisfied for most of the example models of Section 18.2 under the usual assumptions entertained for these models. It is possible to derive recursions similar to those given below under more general conditions. If the normality assumption is dropped, the recursions given below can still be justified. We will return to this issue after having presented them.

The Recursions

We will use the following additional notation in stating the Kalman filter recursions:

$$
\begin{aligned}
&z_{t|s} := E(z_t|y_1, \ldots, y_s),\\
&\Sigma_z(t|s) := \mathrm{Cov}(z_t|y_1, \ldots, y_s),\\
&y_{t|s} := E(y_t|y_1, \ldots, y_s),\\
&\Sigma_y(t|s) := \mathrm{Cov}(y_t|y_1, \ldots, y_s),\\
&(z|y) \sim \mathcal{N}(\mu, \Sigma) \quad \text{means that the conditional distribution of } z \text{ given } y \text{ is multivariate normal with mean } \mu \text{ and covariance matrix } \Sigma.
\end{aligned}
\tag{18.3.3}
$$

Under the previously stated conditions, the normality assumption implies

$$(z_t|y_1, \ldots, y_{t-1}) \sim \mathcal{N}(z_{t|t-1}, \Sigma_z(t|t-1)) \quad \text{for } t = 2, \ldots, T, \tag{18.3.4}$$

$$(z_t|y_1, \ldots, y_t) \sim \mathcal{N}(z_{t|t}, \Sigma_z(t|t)) \quad \text{for } t = 1, \ldots, T, \tag{18.3.5}$$

$$(y_t|y_1, \ldots, y_{t-1}) \sim \mathcal{N}(y_{t|t-1}, \Sigma_y(t|t-1)) \quad \text{for } t = 2, \ldots, T, \tag{18.3.6}$$

and

$$(z_t|y_1, \ldots, y_T) \sim \mathcal{N}(z_{t|T}, \Sigma_z(t|T)), \tag{18.3.7}$$

$$(y_t|y_1, \ldots, y_T) \sim \mathcal{N}(y_{t|T}, \Sigma_y(t|T)) \quad \text{for } t > T. \tag{18.3.8}$$

The conditional means and covariance matrices can be obtained by the following *Kalman filter recursions* which are graphically depicted in Figure 18.1:

Initialization: $z_{0|0} := \mu_0,\ \Sigma_z(0|0) := \Sigma_0$.

Prediction step $(1 \le t \le T)$:

$$
\begin{aligned}
&z_{t|t-1} = \mathbf{B} z_{t-1|t-1} + \mathbf{F} x_{t-1},\\
&\Sigma_z(t|t-1) = \mathbf{B}\Sigma_z(t-1|t-1)\mathbf{B}' + \Sigma_w,\\
&y_{t|t-1} = \mathbf{H}_t z_{t|t-1} + \mathbf{G} x_t,\\
&\Sigma_y(t|t-1) = \mathbf{H}_t \Sigma_z(t|t-1)\mathbf{H}_t' + \Sigma_v.
\end{aligned}
$$

Correction step $(1 \le t \le T)$:

$$
\begin{aligned}
&z_{t|t} = z_{t|t-1} + \mathbf{P}_t(y_t - y_{t|t-1}),\\
&\Sigma_z(t|t) = \Sigma_z(t|t-1) - \mathbf{P}_t \Sigma_y(t|t-1)\mathbf{P}_t',
\end{aligned}
$$

where

$$\mathbf{P}_t := \Sigma_z(t|t-1)\mathbf{H}_t'\Sigma_y(t|t-1)^{-1} \quad (Kalman\ filter\ gain).$$

Although the output variables we have in mind have nonsingular distributions, it may be worth noting that if the inverse of $\Sigma_y(t|t-1)$ does not exist, it may be replaced by a suitable generalized inverse. The recursions proceed by performing the prediction step for $t = 1$. Then the correction step is carried out for $t = 1$. Then the prediction and correction steps are repeated for $t = 2$, and so on.

Forecasting step ($t > T$):

$$z_{t|T} = \mathbf{B}z_{t-1|T} + \mathbf{F}x_{t-1},$$

$$\Sigma_z(t|T) = \mathbf{B}\Sigma_z(t-1|T)\mathbf{B}' + \Sigma_w,$$

$$y_{t|T} = \mathbf{H}_t z_{t|T} + \mathbf{G}x_t,$$

$$\Sigma_y(t|T) = \mathbf{H}_t\Sigma_z(t|T)\mathbf{H}_t' + \Sigma_v.$$

The forecasting step may be carried out recursively for $t = T+1, T+2, \ldots$.

Computational Aspects and Extensions

In practice, in running through the Kalman filter recursions, computational inaccuracies may accumulate in such a way that the actually computed covariance matrices are not positive semidefinite. These and other computational issues were discussed in Anderson & Moore (1979, Chapter 6) and numerical modifications of the recursions were suggested that may help to overcome the possible difficulties (see also Schneider (1992)).

As mentioned previously, it is possible to justify the Kalman filter recursions even if the initial state and the white noise processes are not Gaussian. In that case, the quantities obtained by the recursions are no longer moments of conditional normal distributions, however. For other interpretations of the quantities see, for example, Schneider (1988).

Sometimes reconstruction of the state vectors, given all the information $y_1, \ldots, y_T$, is of interest. For instance, in the random coefficients models of Section 18.2.1, where the state vector z_t contains the coefficients associated with period t, one may want to estimate the states and, hence, the coefficients, given all the sample information $y_1, \ldots, y_T$. We will see a detailed example in Section 18.5. Recursions are also available to compute $z_{t|T}$ and $\Sigma_z(t|T)$ for $t < T$. The evaluation of $z_{t|T}$ for $t < T$ is known as *smoothing*. Under the previous assumptions (including normality),

$$(z_t|y_1, \ldots, y_T) \sim \mathcal{N}(z_{t|T}, \Sigma_z(t|T))$$

for $t = 0, 1, \ldots, T$. The conditional moments may be obtained recursively, starting at the end of the sample and moving backwards, that is, the recursions proceed for $t = T-1, T-2, \ldots, 0$ as follows.

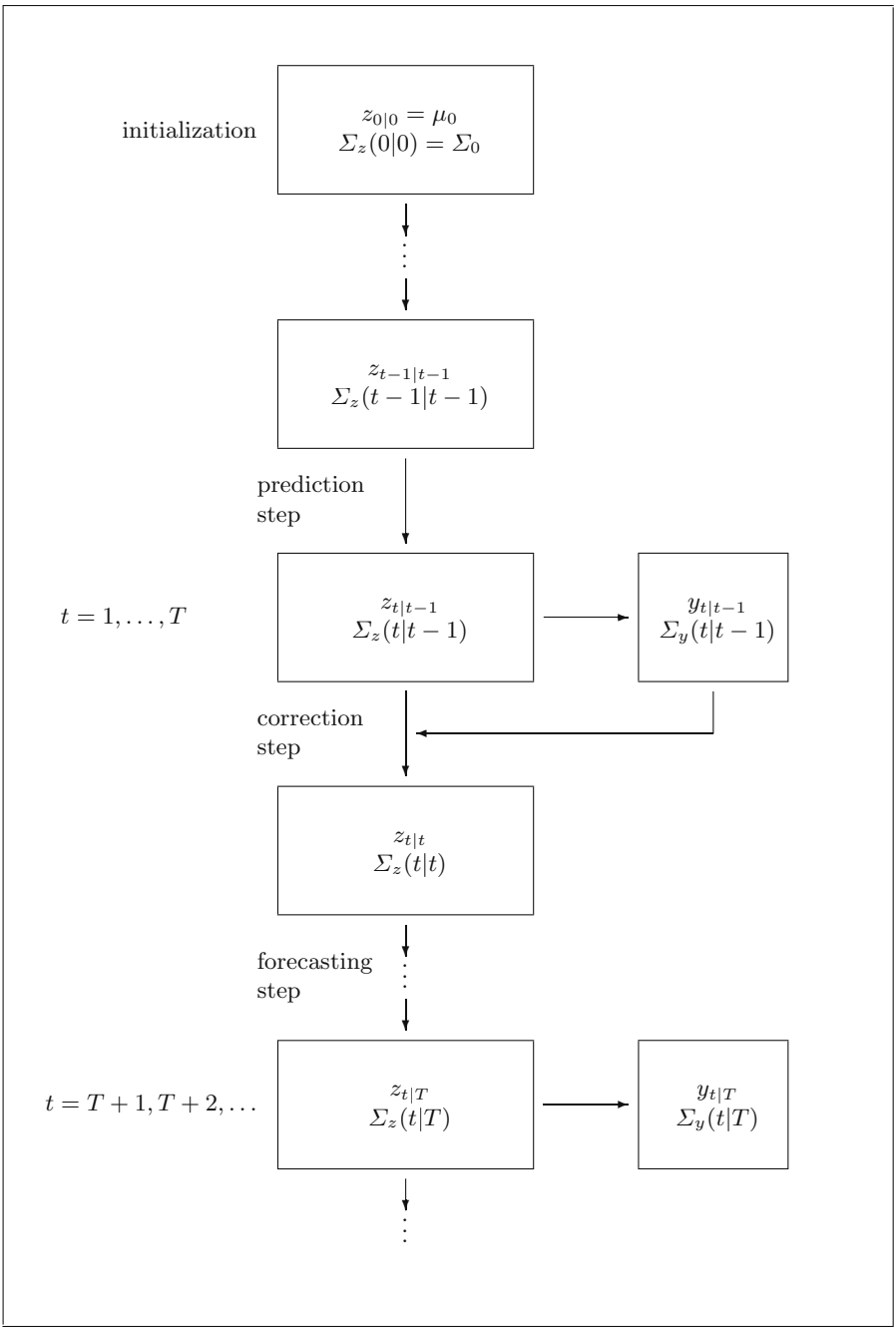

Fig. 18.1. Kalman filter recursions.

Smoothing step ($t < T$):

$$z_{t|T} = z_{t|t} + \mathbf{S}_t(z_{t+1|T} - z_{t+1|t}),$$

$$\Sigma_z(t|T) = \Sigma_z(t|t) - \mathbf{S}_t[\Sigma_z(t+1|t) - \Sigma_z(t+1|T)]\mathbf{S}_t',$$

where

$$\mathbf{S}_t := \Sigma_z(t|t)\mathbf{B}'\Sigma_z(t+1|t)^{-1} \quad (\textit{Kalman smoothing matrix}),$$

(see Anderson & Moore (1979)).

18.3.2 Proof of the Kalman Filter Recursions

The proof follows Anderson & Moore (1979, pp. 39-41) and Meinhold & Singpurwalla (1983). It may be skipped without loss of continuity. We proceed inductively and we use the following properties of multivariate normal distributions (see Propositions B.1 and B.2 of Appendix B):

$$y \sim \mathcal{N}(\mu_y, \Sigma_y), \quad z \sim \mathcal{N}(\mu_z, \Sigma_z) \text{ are independent } (K \times 1) \text{ random vectors}$$
$$\Rightarrow y + z \sim \mathcal{N}(\mu_y + \mu_z, \Sigma_y + \Sigma_z). \tag{18.3.9}$$

If A is a fixed, nonrandom matrix and c a fixed vector,

$$y \sim \mathcal{N}(\mu_y, \Sigma_y) \Rightarrow Ay + c \sim \mathcal{N}(A\mu_y + c, A\Sigma_y A'). \tag{18.3.10}$$

Moreover,

$$\begin{bmatrix} z \\ y \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} \mu_z \\ \mu_y \end{bmatrix}, \begin{bmatrix} \Sigma_z & \Sigma_{zy} \\ \Sigma_{yz} & \Sigma_y \end{bmatrix}\right)$$
$$\Rightarrow (z|y) \sim \mathcal{N}(\mu_z + \Sigma_{zy}\Sigma_y^{-1}(y - \mu_y),\ \Sigma_z - \Sigma_{zy}\Sigma_y^{-1}\Sigma_{yz}). \tag{18.3.11}$$

Here Σ_y^{-1} may be replaced by a generalized inverse, if Σ_y is singular.

We will now demonstrate the prediction and correction steps for $t = 1$. With that goal in mind, we note that by (18.3.9) and (18.3.10) and the joint normality of w_0 and v_1, the two vectors z_1 and y_1 are jointly normally distributed,

$$\begin{aligned}\begin{bmatrix} z_1 \\ y_1 \end{bmatrix} &= \begin{bmatrix} I \\ \mathbf{H}_1 \end{bmatrix} z_1 + \begin{bmatrix} 0 \\ \mathbf{G}x_1 \end{bmatrix} + \begin{bmatrix} 0 \\ I \end{bmatrix} v_1 \\ &= \begin{bmatrix} I \\ \mathbf{H}_1 \end{bmatrix} \mathbf{B}z_0 + \begin{bmatrix} \mathbf{F}x_0 \\ \mathbf{G}x_1 + \mathbf{H}_1\mathbf{F}x_0 \end{bmatrix} + \begin{bmatrix} I & 0 \\ \mathbf{H}_1 & I \end{bmatrix} \begin{bmatrix} w_0 \\ v_1 \end{bmatrix}\end{aligned}$$

$$\sim \mathcal{N}\left(\begin{bmatrix} \mathbf{B}\mu_0 + \mathbf{F}x_0 \\ \mathbf{H}_1(\mathbf{B}\mu_0 + \mathbf{F}x_0) + \mathbf{G}x_1 \end{bmatrix}, \right.$$
$$\left. \begin{bmatrix} I & 0 \\ \mathbf{H}_1 & I \end{bmatrix} \begin{bmatrix} \Sigma_w & 0 \\ 0 & \Sigma_v \end{bmatrix} \begin{bmatrix} I & \mathbf{H}_1' \\ 0 & I \end{bmatrix} + \begin{bmatrix} \mathbf{B} \\ \mathbf{H}_1\mathbf{B} \end{bmatrix} \Sigma_0[\mathbf{B}', \mathbf{B}'\mathbf{H}_1'] \right).$$

Hence,

$$
\begin{aligned}
&z_{1|0} := E(z_1) = \mathbf{B}\mu_0 + \mathbf{F}x_0 = \mathbf{B}z_{0|0} + \mathbf{F}x_0,\\
&\Sigma_z(1|0) := \mathrm{Cov}(z_1) = \Sigma_w + \mathbf{B}\Sigma_0\mathbf{B}' = \mathbf{B}\Sigma_z(0|0)\mathbf{B}' + \Sigma_w,\\
&y_{1|0} := E(y_1) = \mathbf{H}_1 z_{1|0} + \mathbf{G}x_1,\\
&\begin{aligned}\Sigma_y(1|0) := \mathrm{Cov}(y_1) &= \mathbf{H}_1\Sigma_w\mathbf{H}_1' + \Sigma_v + \mathbf{H}_1\mathbf{B}\Sigma_0\mathbf{B}'\mathbf{H}_1'\\ &= \mathbf{H}_1\Sigma_z(1|0)\mathbf{H}_1' + \Sigma_v,\end{aligned}
\end{aligned}
$$

which proves the prediction step for $t = 1$. Using these results and (18.3.11), the conditional distribution of z_1 given y_1 is seen to be

$$
\begin{aligned}
(z_1|y_1) \sim \mathcal{N}[&z_{1|0} + \Sigma_z(1|0)\mathbf{H}_1'\Sigma_y(1|0)^{-1}(y_1 - y_{1|0}),\\
&\Sigma_z(1|0) - \Sigma_z(1|0)\mathbf{H}_1'\Sigma_y(1|0)^{-1}\mathbf{H}_1\Sigma_z(1|0)],
\end{aligned}
$$

which proves the correction step for $t = 1$.

Now the prediction and correction steps can be shown by induction. Suppose the normal distributions in (18.3.4)–(18.3.6) and the prediction and correction steps are correct for $t-1$. Then, using the transition and measurement equations, z_t and y_t have a joint normal distribution

$$
\begin{aligned}
\begin{bmatrix} z_t \\ y_t \end{bmatrix} &= \begin{bmatrix} I \\ \mathbf{H}_t \end{bmatrix} z_t + \begin{bmatrix} 0 \\ \mathbf{G} \end{bmatrix} x_t + \begin{bmatrix} 0 \\ I \end{bmatrix} v_t\\
&= \begin{bmatrix} I \\ \mathbf{H}_t \end{bmatrix} (\mathbf{B}z_{t-1} + \mathbf{F}x_{t-1} + w_{t-1}) + \begin{bmatrix} 0 \\ \mathbf{G} \end{bmatrix} x_t + \begin{bmatrix} 0 \\ I \end{bmatrix} v_t.
\end{aligned}
$$

By the induction assumption and (18.3.9)/(18.3.10), this term has the following conditional normal distribution, given $y_1, \ldots, y_{t-1}$:

$$
\begin{aligned}
\mathcal{N}\Bigg(&\begin{bmatrix} \mathbf{B}z_{t-1|t-1} + \mathbf{F}x_{t-1} \\ \mathbf{H}_t(\mathbf{B}z_{t-1|t-1} + \mathbf{F}x_{t-1}) + \mathbf{G}x_t \end{bmatrix},\\
&\begin{bmatrix} \Sigma_z(t|t-1) & \bullet \\ \mathbf{H}_t\Sigma_z(t|t-1) & \mathbf{H}_t\Sigma_z(t|t-1)\mathbf{H}_t' + \Sigma_v \end{bmatrix}\Bigg),
\end{aligned} \tag{18.3.12}
$$

where $\Sigma_z(t|t-1) = \mathbf{B}\Sigma_z(t-1|t-1)\mathbf{B}' + \Sigma_w$. This proves the prediction step. Application of (18.3.11) to (18.3.12) gives the conditional distribution of z_t given $y_1, \ldots, y_t$ and proves the correction step.

It remains to prove the forecasting step. Again by induction $(z_t|y_1, \ldots, y_T)$ and $(y_t|y_1, \ldots, y_T)$ both have normal distributions with the first and second moments as stated in the forecasting step.

18.4 Maximum Likelihood Estimation of State Space Models

In this section, we consider ML estimation of the state space system given in Section 18.3.1. We assume that the matrices $\mathbf{B}$, $\mathbf{F}$, $\mathbf{H}_t$, $\mathbf{G}$, Σ_w, Σ_v, Σ_0, and

the vector μ_0 depend on a vector of time invariant parameters $\boldsymbol{\delta}$. In other words, $\boldsymbol{\delta}$ is time invariant, even if $\mathbf{H}_t$ is not. For a given $\boldsymbol{\delta}$, the matrices are assumed to be uniquely determined and at least twice continuously differentiable with respect to the elements of $\boldsymbol{\delta}$. For instance, in the state space model (18.2.8)/(18.2.9) which represents the finite order VAR process (18.2.6),

$$\boldsymbol{\delta} = \begin{bmatrix} \text{vec}[\nu, A_1, \ldots, A_p] \\ \text{vech}(\Sigma_u) \end{bmatrix},$$

if no constraints are placed on the VAR coefficients or Σ_u and if the initial conditions $y_{-p+1}, \ldots, y_0$ are assumed to be known and fixed. The objective in this section is to estimate $\boldsymbol{\delta}$. We will set up the log-likelihood function first. Then we discuss its maximization and, finally, the asymptotic properties of the ML estimators are considered.

18.4.1 The Log-Likelihood Function

By Bayes' theorem, the sample density function can be written as

$$\begin{aligned} f(y_1, \ldots, y_T; \boldsymbol{\delta}) &= f(y_1; \boldsymbol{\delta}) f(y_2, \ldots, y_T | y_1; \boldsymbol{\delta}) \\ &\vdots \\ &= f(y_1; \boldsymbol{\delta}) f(y_2 | y_1; \boldsymbol{\delta}) \cdots f(y_T | y_1, \ldots, y_{T-1}; \boldsymbol{\delta}). \end{aligned}$$

Thus, using the notation of the previous section and assuming that y_t has dimension K, the Gaussian log-likelihood for the present case is

$$\begin{aligned} \ln l(\boldsymbol{\delta} | y_1, \ldots, y_T) &= \ln f(y_1, \ldots, y_T; \boldsymbol{\delta}) \\ &= \ln f(y_1; \boldsymbol{\delta}) + \sum_{t=2}^{T} \ln f(y_t | y_1, \ldots, y_{t-1}; \boldsymbol{\delta}) \\ &= -\frac{KT}{2} \ln(2\pi) - \frac{1}{2} \sum_{t=1}^{T} \ln |\Sigma_y(t|t-1)| \\ &\quad -\frac{1}{2} \sum_{t=1}^{T} (y_t - y_{t|t-1})' \Sigma_y(t|t-1)^{-1} (y_t - y_{t|t-1}), \end{aligned} \tag{18.4.1}$$

where we have used that $y_{1|0} := E(y_1)$, $\Sigma_y(1|0) := \text{Cov}(y_1)$, and

$$(y_t | y_1, \ldots, y_{t-1}) \sim \mathcal{N}(y_{t|t-1}, \Sigma_y(t|t-1)), \quad t = 1, \ldots, T,$$

from Section 18.3.1. Here both $y_{t|t-1}$ and $\Sigma_y(t|t-1)$ depend in general on the parameter vector $\boldsymbol{\delta}$. If a specific vector $\boldsymbol{\delta}$ is given, all the quantities in the log-likelihood function can be computed with the Kalman filter recursions. Thus, the Kalman filter is seen to be a useful tool for evaluating the log-likelihood

function of a wide range of models. Note also that we have considered likelihood approximations for VARMA processes in Chapter 12. In the present framework, the exact likelihood may be obtained (see also Solo (1984)).

To simplify the expression for the log-likelihood given in (18.4.1), we use the following notation:

$$e_t(\boldsymbol{\delta}) := y_t - y_{t|t-1} \quad \text{and } \Sigma_t(\boldsymbol{\delta}) := \Sigma_y(t|t-1). \tag{18.4.2}$$

This notation makes the dependence on $\boldsymbol{\delta}$ explicit. Occasionally, we will, however, drop $\boldsymbol{\delta}$. With this notation, the log-likelihood function can be written as

$$\ln l(\boldsymbol{\delta}) = -\frac{KT}{2}\ln(2\pi) - \frac{1}{2}\sum_{t=1}^{T}[\ln|\Sigma_t(\boldsymbol{\delta})| + e_t(\boldsymbol{\delta})'\Sigma_t(\boldsymbol{\delta})^{-1}e_t(\boldsymbol{\delta})]. \tag{18.4.3}$$

18.4.2 The Identification Problem

Recall from the discussion in Chapter 12 that unique maximization of the likelihood function and asymptotic inference require an identified or unique parameterization. Identification is not automatic in the present context because, for instance, VARMA models are not identified without specific restrictions and VARMA processes are just special cases of the presently considered models. Hence, the identification or uniqueness problem is inherent in the general linear state space model, too. We will state the problem here again in sufficient generality to cover the present case.

Let $\mathbf{y} := \text{vec}(y_1, \ldots, y_T)$ be the vector of observed random variables and denote its distribution by $F(\mathbf{y}; \boldsymbol{\delta}_0)$, where $\boldsymbol{\delta}_0$ is the *true parameter vector*. We assume that the true distribution of $\mathbf{y}$ is a member of the parametric family

$$\{F(\mathbf{y}; \boldsymbol{\delta}) | \boldsymbol{\delta} \in \mathbb{D}\},$$

where $\mathbb{D} \subset \mathbb{R}^n$ is the parameter space. The vector $\boldsymbol{\delta}_0$ is said to be *identified* or *identifiable* if it is the only vector in $\mathbb{D}$ which gives rise to the distribution of $\mathbf{y}$. In other words, for any $\boldsymbol{\delta}_1 \in \mathbb{D}$,

$$\boldsymbol{\delta}_1 \neq \boldsymbol{\delta}_0 \Rightarrow F(\mathbf{y}; \boldsymbol{\delta}_1) \neq F(\mathbf{y}; \boldsymbol{\delta}_0) \quad \text{(for at least one } \mathbf{y}). \tag{18.4.4}$$

To compute ML estimators and to derive asymptotic properties it is actually sufficient that $\boldsymbol{\delta}_0$ has a neighborhood in which it is uniquely determined by the true distribution of $\mathbf{y}$. To distinguish this case from one where uniqueness follows for the whole parameter space, the vector $\boldsymbol{\delta}_0$ or the model is often called *locally identified* or *locally identifiable* if there exists a neighborhood $\mathbb{U}(\boldsymbol{\delta}_0)$ of $\boldsymbol{\delta}_0$ such that (18.4.4) holds for any $\boldsymbol{\delta}_1 \in \mathbb{U}(\boldsymbol{\delta}_0)$. In contrast, the model or parameter vector is *globally identifiable* or *globally identified* if (18.4.4) holds for all $\boldsymbol{\delta}_1 \in \mathbb{D}$.

Because the negative log-likelihood function has a locally unique minimum if its Hessian matrix is positive definite, identification conditions for state space models may be formulated via the information matrix. If we are interested in *asymptotic* properties of estimators, it is sufficient to obtain identification in large samples. Hence, under some regularity conditions, the identification assumption may be disguised in the requirement of a positive definite asymptotic information matrix. In a later proposition giving asymptotic properties of the ML estimators, to ensure identification, we will include the condition that the sequence of normalized information matrices, $\mathcal{I}(\boldsymbol{\delta}_0)/T$, is bounded from below by a positive definite matrix, as T goes to infinity. In special case models, other identification conditions are often easier to deal with and are therefore preferred. For example, for VARMA processes the identification conditions given in Section 12.1.2 may be used.

18.4.3 Maximization of the Log-Likelihood Function

From some previous chapters we know that maximization of the log-likelihood function is in general a nonlinear optimization problem. Therefore, numerical methods are required for its solution. One possibility is a gradient algorithm as described, for example, in Section 12.3.2 for iteratively minimizing $-\ln l$. Recall that the general form of the i-th iteration step is

$$\boldsymbol{\delta}_{i+1} = \boldsymbol{\delta}_i - s_i D_i \left[\frac{\partial(-\ln l)}{\partial \boldsymbol{\delta}} \Big|_{\boldsymbol{\delta}_i} \right], \tag{18.4.5}$$

where s_i is the step length and D_i is a positive definite direction matrix. The inverse information matrix is one possible choice for this matrix. In that case, the method is called *scoring algorithm*. We will provide the ingredients for this algorithm in the following, that is, we will give expressions for the gradient of $\ln l$ and an estimator of the information matrix. There are various ways to choose the step length s_i. For instance, it could be chosen so as to optimize the progress towards the minimum. Another alternative would be to simply set $s_i = 1$. We will not discuss the step length selection in further detail here because it is of limited importance for the statistical analysis of the model.

The Gradient of the Log-Likelihood

From (18.4.3), we get

$$\frac{\partial \ln l}{\partial \boldsymbol{\delta}'} = -\frac{1}{2} \sum_{t=1}^{T} \left[\operatorname{vec}\left(\frac{\partial \ln |\Sigma_t|}{\partial \Sigma_t} \right)' \frac{\partial \operatorname{vec}(\Sigma_t)}{\partial \boldsymbol{\delta}'} + \frac{\partial \operatorname{tr}(e_t' \Sigma_t^{-1} e_t)}{\partial \boldsymbol{\delta}'} \right]$$

$$= -\frac{1}{2}\sum_t \left[\operatorname{vec}(\Sigma_t^{-1})' \frac{\partial \operatorname{vec}(\Sigma_t)}{\partial \boldsymbol{\delta}'} + 2e_t'\Sigma_t^{-1}\frac{\partial e_t}{\partial \boldsymbol{\delta}'} \right.$$
$$\left. - \operatorname{vec}(\Sigma_t^{-1}e_t e_t'\Sigma_t^{-1})' \frac{\partial \operatorname{vec}(\Sigma_t)}{\partial \boldsymbol{\delta}'} \right]$$

$$= -\frac{1}{2}\sum_t \left[\operatorname{vec}[\Sigma_t^{-1}(I_K - e_t e_t'\Sigma_t^{-1})]' \frac{\partial \operatorname{vec}(\Sigma_t)}{\partial \boldsymbol{\delta}'} + 2e_t'\Sigma_t^{-1}\frac{\partial e_t}{\partial \boldsymbol{\delta}'} \right], \tag{18.4.6}$$

where $\partial e_t/\partial \boldsymbol{\delta}'$ may be replaced by $-\partial y_{t|t-1}/\partial \boldsymbol{\delta}'$.

The Information Matrix

Using $E(e_t e_t') = \Sigma_t$ $(= \Sigma_y(t|t-1))$ and $E[e_t(\boldsymbol{\delta}_0)] = 0$, straightforward application of the rules for matrix and vector differentiation yields the information matrix,

$$\begin{aligned} \mathcal{I}(\boldsymbol{\delta}_0) &= -E\left[\frac{\partial^2 \ln l}{\partial \boldsymbol{\delta} \partial \boldsymbol{\delta}'}\bigg|_{\boldsymbol{\delta}_0} \right] \\ &= \frac{1}{2}\sum_{t=1}^{T} \left[\frac{\partial \operatorname{vec}(\Sigma_t)'}{\partial \boldsymbol{\delta}} (\Sigma_t^{-1} \otimes \Sigma_t^{-1}) \frac{\partial \operatorname{vec}(\Sigma_t)}{\partial \boldsymbol{\delta}'} \right. \\ &\qquad \left. + 2E\left(\frac{\partial e_t'}{\partial \boldsymbol{\delta}} \Sigma_t^{-1} \frac{\partial e_t}{\partial \boldsymbol{\delta}'} \right) \right]. \end{aligned} \tag{18.4.7}$$

Because the true parameter values involved in this expression are unknown, they are replaced by estimators and the expectation is simply dropped. For instance, in the i-th iteration of the scoring algorithm, $\boldsymbol{\delta}_i$ is used as an estimator for $\boldsymbol{\delta}_0$.

Discussion of the Scoring Algorithm

The scoring algorithm may have poor convergence properties far away from the maximum of the log-likelihood function. On the other hand, it has very good convergence properties close to the maximum. Unfortunately, it may be expensive in terms of computation time because it requires (possibly numerical) evaluation of derivatives in each iteration. Therefore, other maximization methods were proposed in the literature. Notably the EM (expectation step-maximization step) algorithm of Dempster, Laird & Rubin (1977) was found to be useful in practice (see Watson & Engle (1983), Schneider (1992)). The EM algorithm is an iterative algorithm which has the advantage of involving much cheaper computations in each iteration step than the scoring algorithm. On the other hand, convergence of the former is slower than that of the latter algorithm. Nicholls & Pagan (1985) and Schneider (1991, 1992) suggested

combining the EM and the scoring algorithms. This proposal may be useful if no good initial estimator $\boldsymbol{\delta}_1$ is available from where to start the scoring algorithm. Another alternative is to use the so-called subspace algorithm for getting initial values (e.g., Bauer & Wagner (2002)).

18.4.4 Asymptotic Properties of the ML Estimator

We consider the state space model from Section 18.3.1 with transition equation

$$z_t = \mathbf{B}z_{t-1} + \mathbf{F}x_{t-1} + w_{t-1} \tag{18.4.8}$$

and measurement equation

$$y_t = \mathbf{H}_t z_t + \mathbf{G}x_t + v_t. \tag{18.4.9}$$

All assumptions of Section 18.3.1 are taken to be satisfied. In addition we assume that

(i) the true parameter vector $\boldsymbol{\delta}_0$ is in the interior of the parameter space which is supposed to be compact;
(ii) $\mathbf{H}_t = (x_t \otimes I)J$, where J is a known selection matrix such as $J = [I_K : 0 : \cdots : 0]$ or $\mathbf{H}_t = \mathbf{H}$ is a time invariant nonstochastic matrix;
(iii) the inputs x_t are nonstochastic and uniformly bounded, that is, there exist real numbers c_1 and c_2 such that $c_1 \leq x_t' x_t \leq c_2$ for all $t = 0, 1, 2, \ldots$;
(iv) the sequence of normalized information matrices is bounded from below by a positive definite matrix, that is, there exists a constant c such that $T^{-1}\mathcal{I}(\boldsymbol{\delta}_0) > cI_n$ or, in other words, $T^{-1}\mathcal{I}(\boldsymbol{\delta}_0) - cI_n$ is positive definite, as $T \to \infty$;
(v) all eigenvalues of $\mathbf{B}$ have modulus less than 1.

As we have discussed in Section 18.4.2, (iv) is an identification condition. The last assumption is a stability condition, and (iii) guarantees that the input variables have no trends. We have seen in Chapter 7 that the standard asymptotic theory may not apply for trending variables. Therefore, they are excluded here. With these assumptions, the following proposition can be established.

Proposition 18.1 (*Asymptotic Properties of the ML Estimator*)
With all the assumptions stated in the foregoing, the ML estimator $\widetilde{\boldsymbol{\delta}}$ of $\boldsymbol{\delta}_0$ is consistent and asymptotically normally distributed,

$$\sqrt{T}(\widetilde{\boldsymbol{\delta}} - \boldsymbol{\delta}_0) \xrightarrow{d} \mathcal{N}(0, \Sigma_{\widetilde{\boldsymbol{\delta}}}), \tag{18.4.10}$$

where

$$\Sigma_{\widetilde{\boldsymbol{\delta}}} = \lim T\mathcal{I}(\boldsymbol{\delta}_0)^{-1}$$

is the inverse asymptotic information matrix. It is consistently estimated by substituting the ML estimators for unknown parameters in (18.4.7), dropping the expectation operator and dividing by T. ■

Pagan (1980) gives a proof of this proposition based on Crowder (1976) (see also Schneider (1988)). Other sets of conditions are possible to accommodate the situation where the inputs x_t are stochastic. They may, in fact, contain lagged y_t's. Moreover, $\mathbf{B}$ may have eigenvalues on the unit circle if it does not contain unknown parameters. The reader is referred to the articles by Pagan (1980), Nicholls & Pagan (1985), Schneider (1988), and to a book by Caines (1988) for details. When particular models are considered, different sets of assumptions are often preferable for two reasons. First, other sets of conditions may be easier to verify or to understand for special models. Second, the conditions of Proposition 18.1 or the modifications mentioned in the foregoing may not be satisfied. We will see an example of the latter case shortly.

A number of alternatives to ML estimation were suggested, see, e.g., Anderson & Moore (1979), Nicholls & Pagan (1985), Schneider (1988), and Bauer & Wagner (2002) for more details and references.

It may be worth noting that application of the Kalman filter to systems with estimated parameters produces state estimates and precision matrices that do not take into account the estimation variability. Watanabe (1985) and Hamilton (1986) considered the properties of state estimators obtained with estimated parameter Kalman filter recursions. Furthermore, a state space framework for unit root processes was presented by Bauer & Wagner (2003).

18.5 A Real Data Example

As an illustrative example, we consider a dynamic consumption function with time varying coefficients,

$$\begin{aligned} y_t &= \gamma_{0t} + \gamma_{1t}x_t + \gamma_{2t}x_{t-1} + \gamma_{3t}y_{t-1} + \gamma_{4t}x_{t-2} + \gamma_{5t}y_{t-2} + v_t \\ &= X_t'\boldsymbol{\gamma}_t + v_t, \end{aligned} \tag{18.5.1}$$

where

$$X_t := \begin{bmatrix} 1 \\ x_t \\ x_{t-1} \\ y_{t-1} \\ x_{t-2} \\ y_{t-2} \end{bmatrix} \quad \text{and} \quad \boldsymbol{\gamma}_t := \begin{bmatrix} \gamma_{0t} \\ \gamma_{1t} \\ \gamma_{2t} \\ \gamma_{3t} \\ \gamma_{4t} \\ \gamma_{5t} \end{bmatrix}.$$

Here y_t and x_t represent rates of change (first differences of logarithms) of consumption and income, respectively. Suppose that the coefficient vector $\boldsymbol{\gamma}_t$ differs from $\boldsymbol{\gamma}_{t-1}$ by an additive random disturbance, that is,

$$\boldsymbol{\gamma}_t = \boldsymbol{\gamma}_{t-1} + w_{t-1}. \tag{18.5.2}$$

In other words, $\boldsymbol{\gamma}_t$ is driven by a (multivariate) random walk. Clearly, (18.5.1) and (18.5.2) represent the measurement and transition equations of a state

space model with $\mathbf{H}_t = X_t'$ and $\mathbf{B} = I_6$. We complete the model by assuming that v_t and w_t are independent Gaussian white noise processes, $v_t \sim \mathcal{N}(0, \sigma_v^2)$ and $w_t \sim \mathcal{N}(0, \Sigma_w)$, where

$$\Sigma_w = \begin{bmatrix} \sigma_{w_0}^2 & & 0 \\ & \ddots & \\ 0 & & \sigma_{w_5}^2 \end{bmatrix} \tag{18.5.3}$$

is a diagonal matrix. Furthermore, the initial state $\boldsymbol{\gamma}_0$ is also assumed to be normally distributed, $\boldsymbol{\gamma}_0 \sim \mathcal{N}(\overline{\boldsymbol{\gamma}}_0, \Sigma_0)$, and independent of v_t and w_t. Admittedly, our assumed model is quite simple. Still, it is useful to illustrate some concepts considered in the previous sections.

Assuming that a sample $\mathbf{y} = (y_1, \ldots, y_T)'$ is available, the log-likelihood function of our model is

$$\begin{aligned} &\ln l(\sigma_v^2, \Sigma_w, \overline{\boldsymbol{\gamma}}_0, \Sigma_0 | \mathbf{y}) \\ &= -\frac{T}{2}\ln(2\pi) - \frac{1}{2}\sum_{t=1}^{T} \ln|\Sigma_y(t|t-1)| - \frac{1}{2}\sum_{t=1}^{T}(y_t - y_{t|t-1})^2/\Sigma_y(t|t-1), \end{aligned} \tag{18.5.4}$$

where $\Sigma_y(t|t-1)$ is a scalar ((1×1) matrix) because y_t is a univariate variable. The log-likelihood function may be evaluated with the Kalman filter recursions for given parameters $\sigma_v^2, \Sigma_w, \overline{\boldsymbol{\gamma}}_0$, and Σ_0. The maximization problem may be solved with an iterative algorithm. Once estimates of the parameters σ_v^2 and Σ_w are available, estimates $\boldsymbol{\gamma}_{t|T}$ of the coefficients of the consumption function (18.5.1) may be obtained with the smoothing recursions given in Section 18.3.1.

Using first differences of logarithms of the quarterly consumption and income data given in File E1 for the years 1960 to 1982, we have estimated the parameters of the state space model (18.5.1)/(18.5.2). The ML estimates of the parameters of interest, namely the variances σ_v^2 and $\sigma_{w_i}^2$, $i = 0, 1, \ldots, 5$, together with estimated standard errors (square roots of the diagonal elements of the estimated inverse information matrix) and corresponding t-ratios are given in Table 18.1.

The interpretation of the standard errors and t-ratios needs caution for various reasons. In Proposition 18.1, where the asymptotic distribution of the ML estimators is given, we have assumed that all eigenvalues of the transition matrix $\mathbf{B}$ have modulus less than 1. This condition is clearly not satisfied in the present example, where $\mathbf{B} = I_6$ and, thus, all six eigenvalues are equal to 1. However, as mentioned in Section 18.4.4, the condition on the eigenvalues of $\mathbf{B}$ is not crucial if $\mathbf{B}$ is a known matrix which does not contain unknown parameters. Of course, setting $\mathbf{B} = I_6$ is just an assumption which may or may not be adequate.

A further deviation from the assumptions of Proposition 18.1 is that the inputs X_t contain lagged endogenous variables and hence are stochastic. Again,

Table 18.1. ML estimates for the example model

parameter	estimate	standard error	t-ratio
σ_v^2	3.91×10^{-5}	1.99×10^{-5}	1.97
$\sigma_{w_0}^2$	2.04×10^{-5}	2.33×10^{-5}	.88
$\sigma_{w_1}^2$	$.14 \times 10^{-2}$	1.08×10^{-2}	.13
$\sigma_{w_2}^2$	$.46 \times 10^{-2}$	$.92 \times 10^{-2}$	.50
$\sigma_{w_3}^2$	$.45 \times 10^{-2}$	1.11×10^{-2}	.41
$\sigma_{w_4}^2$	$.51 \times 10^{-2}$	$.94 \times 10^{-2}$	.54
$\sigma_{w_5}^2$	$.62 \times 10^{-2}$	1.16×10^{-2}	.54

we have mentioned in Section 18.4.4 that this assumption is not necessarily critical. The conditions of Proposition 18.1 could be modified so as to allow for lagged dependent variables.

Another assumption that may be problematic is the normality of the white noise sequences and the initial state. The normality assumption may be checked by computing the skewness and kurtosis of the standardized quantities $(y_t - y_{t|t-1})/\Sigma_y(t|t-1)^{1/2}$. A test for nonnormality may then be based on the χ^2-statistic involving both skewness and kurtosis as described in Chapter 4, Section 4.5. For the present example, the statistic assumes the value 3.00 and has a $\chi^2(2)$-distribution under the null hypothesis of normality. Thus, it is not significant at any conventional level.

Finally, we have assumed in Proposition 18.1 that the true parameter values lie in the interior of the parameter space. Given that the variance estimates are quite small compared to their estimated standard errors, it is possible that at least the $\sigma_{w_i}^2$ are in fact zero and, thus, lie on the boundary of the feasible parameter space. If the $\sigma_{w_i}^2$ are actually zero, the $\boldsymbol{\gamma}_t$ are time invariant in our model which would be a hypothesis of considerable interest. It would permit us to work with a constant coefficient specification. Unfortunately, if $\sigma_{w_i}^2 = 0$, the corresponding t-ratio does not have an asymptotic standard normal distribution in general. Thus, we cannot use the t-ratios given in Table 18.1 for testing the null hypotheses $\sigma_{w_i}^2 = 0$, $i = 0, 1, \ldots, 5$.

In the present context, we may ignore the problems related to the asymptotic theory for the moment and simply regard the model as a descriptive tool. Using the estimated values of the parameters of the model, we may consider the smoothing estimates $\boldsymbol{\gamma}_{t|T}$ of the states (the coefficients of the consumption function). They are plotted in Figure 18.2. The two-standard error bounds which are also shown in the figure are computed from the $\Sigma_{\boldsymbol{\gamma}}(t|T)$. These quantities are obtained with the smoothing recursions given in Section 18.3.1. From the plots in Figure 18.2, it can be seen that the intercept term γ_{0t} is the only coefficient that exhibits substantial variation through time. For instance, a considerable downturn is observed in 1966/1967, where the West German economy was in a recession. All the other coefficients show relatively little variation through time, although γ_{3t}, γ_{4t}, and γ_{5t} (the coefficients of

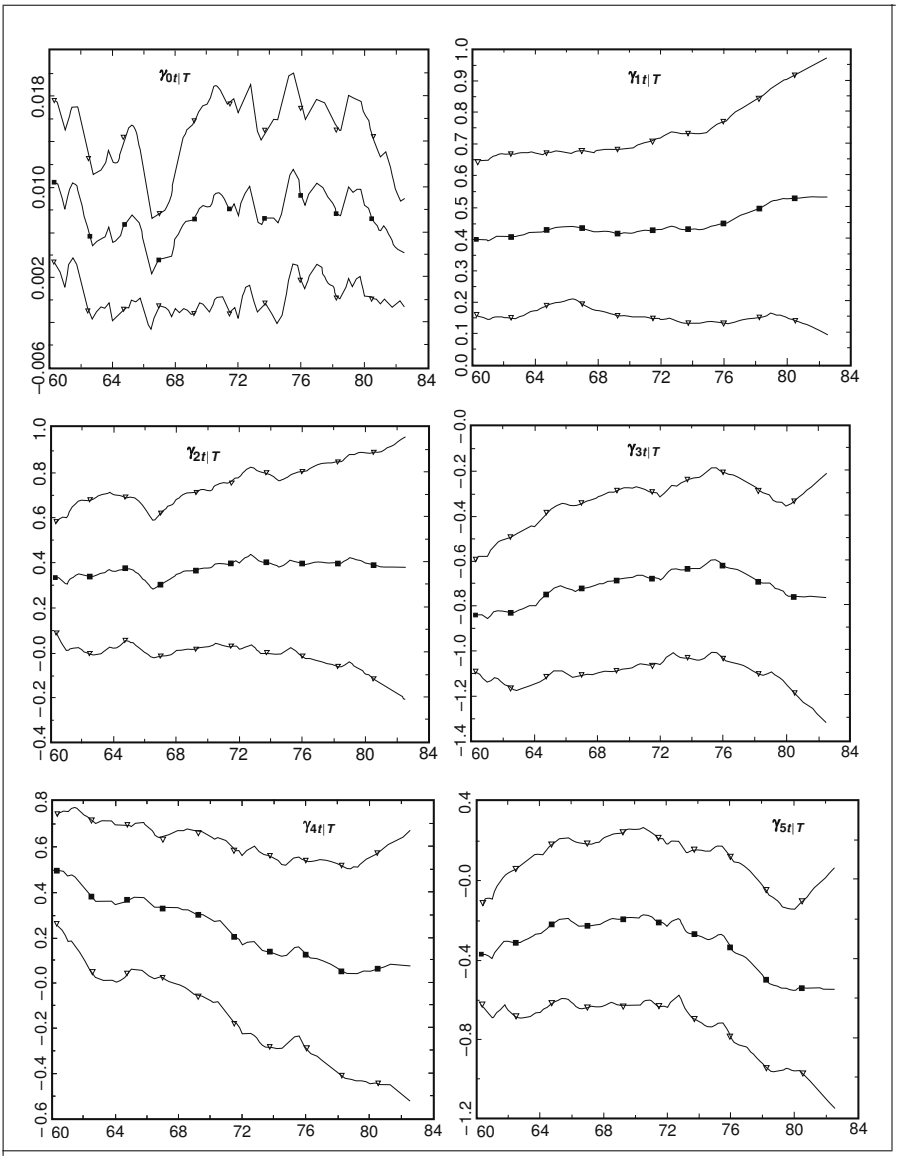

Fig. 18.2. Smoothing estimates of the consumption function coefficients (—■— coefficient estimate, —▽— estimated two-standard error bound).

y_{t-1}, x_{t-2}, and y_{t-2}, respectively) have a tendency to decline in the second half of the 1970s. However, given the estimated two-standard error bounds, overall the results support a specification with constant coefficients of current and lagged income and lagged consumption.

As mentioned previously, this example is quite simplistic. It is just meant to illustrate some of the concepts discussed in this chapter. In a more general model, other lags of income and/or consumption could appear on the right-hand side of the consumption function, the coefficients could be generated by a more general VAR model and the covariance matrix of w_t could be nondiagonal. Moreover, the consumption function may just be a part of a system of equations.

Given that we have discussed different models for the same data in previous chapters, the example also illustrates that there is not just one possible model or model class for the generation process of a multiple time series. The reader may wonder which of the models we have considered in this and the previous chapters is "best". That, however, depends on the questions of interest. In other words, the time series analyst has to decide on the model with the objective of his or her analysis in mind. In this book, we have just tried to introduce some of the possible tools in this venture. With these tools in hand, the analyst is hoped to be able to approach his or her problems of interest in a superior way, with an improved sense of the available possibilities and the potential pitfalls.

18.6 Exercises

Problem 18.1
Write the VARMAX model

$$\begin{aligned} y_t &= A_1 y_{t-1} + \cdots + A_p y_{t-p} + B_0 x_t + \cdots + B_s x_{t-s} \\ &\quad + u_t + M_1 u_{t-1} + \cdots + M_q u_{t-q} \end{aligned}$$

in state space form.

Problem 18.2
Suppose that in the dynamic factor analytic model, $y_t = \mathbf{L} f_t + u_t$, the common factors f_t are generated by the VARMA(p, q) process,

$$f_t = A_1 f_{t-1} + \cdots + A_p f_{t-p} + \eta_t + M_1 \eta_{t-1} + \cdots + M_q \eta_{t-q},$$

and the individual factors u_t are generated by the VARMA(r, s) process

$$u_t = C_1 u_{t-1} + \cdots + C_r u_{t-r} + \varepsilon_t + D_1 \varepsilon_{t-1} + \cdots + D_s \varepsilon_{t-s}.$$

Write the model in state space form.

Problem 18.3
Assume that y_t is generated according to

$$y_t = A_t Y_{t-1} + C_t x_t + v_t,$$

where $A_t := [A_{1t}, \ldots, A_{pt}]$ and $Y_{t-1} := [y'_{t-1}, \ldots, y'_{t-p}]'$. Suppose that $\boldsymbol{\alpha}_t := \text{vec}[A_t : C_t]$ is driven by the VARMA(r, s) process

$$\boldsymbol{\alpha}_t = D_1 \boldsymbol{\alpha}_{t-1} + \cdots + D_r \boldsymbol{\alpha}_{t-r} + \boldsymbol{\eta}_t + M_1 \boldsymbol{\eta}_{t-1} + \cdots + M_s \boldsymbol{\eta}_{t-s}.$$

Write the model in state space form.

Problem 18.4
Write down explicitly the first two steps of the Kalman filter recursions.

Problem 18.5
Suppose the scalar observable variable y_t is generated by the random coefficient regression model

$$y_t = \nu + x_t \beta_t + v_t, \quad t = 1, \ldots, T,$$

where $\beta_t = \alpha \beta_{t-1} + w_t$ is driven by an AR(1) process. Suppose further that v_t and w_t are independent zero mean Gaussian white noise processes with variance 1 and let β_0 be a standard normal random variable.

(a) Determine the conditional distribution of β_t given $y_1, \ldots, y_{t-1}$.
(b) Write down the log-likelihood function of the model and derive its gradient. Find an expression for the information matrix.

Problem 18.6
Consider the K-dimensional Gaussian stable VAR(1) process $y_t = A y_{t-1} + u_t$ with $y_0 \sim \mathcal{N}(0, 0)$ and $u_t \sim \mathcal{N}(0, \Sigma_u)$ for $t = 1, 2, \ldots$. Use the Kalman filter recursions to determine $y_{t|t-1}$.

(a) Show that $y_{t|t-1} = A y_{t-1}$.
(b) Show that the conditions of Proposition 18.1 are satisfied if

$$\boldsymbol{\delta} = \begin{bmatrix} \text{vec}(A) \\ \text{vech}(\Sigma_u) \end{bmatrix}.$$

Problem 18.7
Repeat the analysis of Section 18.5 with the same data and the state space model consisting of the measurement equation

$$y_t = \gamma_{0t} + \gamma_{1t} x_t + \gamma_{2t} y_{t-1} + v_t$$

and the transition equation

$$\begin{bmatrix} \gamma_{0,t} \\ \gamma_{1,t} \\ \gamma_{2,t} \end{bmatrix} = \begin{bmatrix} \gamma_{0,t-1} \\ \gamma_{1,t-1} \\ \gamma_{2,t-1} \end{bmatrix} + w_{t-1}.$$