10

Systems of Dynamic Simultaneous Equations

10.1 Background

This chapter serves to point out some possible extensions of the models considered so far and to draw attention to potential problems related to such extensions. So far, we have assumed that all stochastic variables of a system have essentially the same status in that they are all determined within the system. In other words, the model describes the joint generation process of all the observable variables of interest. In practice, the generation process may be affected by other observable variables which are determined outside the system of interest. Such variables are called *exogenous* or *unmodelled* variables. In contrast, the variables determined within the system are called *endogenous*. Although deterministic terms can be included in the set of unmodelled variables, we often have stochastic variables in mind in this category. For instance, weather related variables such as rainfall or hours of sunshine are usually regarded as stochastic exogenous variables. As another example of the latter type of variables, if a small open economy is being studied, the price level or the output of the rest of the world may be regarded as exogenous. A model which specifies the generation process of some variables conditionally on some other unmodelled variables is sometimes called a *conditional* or *partial model* because it describes the generation process of a subset of the variables only.

A model with unmodelled variables may have the structural form

$$\mathsf{A}y_t = A_1^* y_{t-1} + \cdots + A_p^* y_{t-p} + B_0^* x_t + B_1^* x_{t-1} + \cdots + B_s^* x_{t-s} + w_t, \quad (10.1.1)$$

where $y_t = (y_{1t}, \dots, y_{Kt})'$ is a K-dimensional vector of endogenous variables, $x_t = (x_{1t}, \dots, x_{Mt})'$ is an M-dimensional vector of unmodelled variables, A is $(K \times K)$ and represents the instantaneous relations between the endogenous variables, the A_i^*'s and B_j^*'s are $(K \times K)$ and $(K \times M)$ coefficient matrices, respectively, and w_t, is a K-dimensional error vector. The vector x_t may contain both stochastic and non-stochastic components. For example, it may include intercept terms, seasonal dummies, and the amount of rainfall in a specific region. If the error term w_t is white noise, a model of the type (10.1.1) is

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

sometimes called a VARX(p, s) model in the following. More generally, models of the form (10.1.1) are often called *linear systems* because they are obviously linear in all variables. In the econometrics literature, the label (linear) *dynamic simultaneous equations model* (SEM) is used for such a model. Because we often have systems of economic variables in mind in the following discussion, we will use this name occasionally. We will also consider a vector error correction version of the model which is useful when cointegrated variables are involved.

Other names that are occasionally found in the related literature are *transfer function models* or *distributed lag models*. These terms will become more plausible in the next section, where different representations and some properties of our basic model (10.1.1) will be discussed. Estimation is briefly considered in Section 10.3 and some remarks on model specification and model checking follow in Section 10.4. Possible uses of such models, namely forecasting, multiplier analysis, and control, are treated in Sections 10.5–10.7. Concluding remarks are contained in Section 10.8. It is not the purpose of this chapter to give a detailed and complete account of all these topics. The chapter is just meant to give some guidance to possible extensions of the by now familiar VAR models and VECMs, the related problems and some further reading.

10.2 Systems with Unmodelled Variables

10.2.1 Types of Variables

In the dynamic simultaneous equations model (10.1.1), we have partitioned the observables in two groups, y_t and x_t. The components of y_t are endogenous variables and the components of x_t are the unmodelled or exogenous variables. Although we have given some explanation of the differences between the two groups of variables, we have not given a precise definition of the terms endogenous and exogenous so far. The idea is that the endogenous variables are determined within the system, whereas the unmodelled, exogenous variables are those on which we can condition the analysis without affecting the results of interest. Because there are different possible objectives of an analysis, there are also different notions of exogeneity. For example, if we are interested in estimating a particular parameter vector $\boldsymbol{\gamma}$, say, x_t is exogenous if the estimation properties do not suffer from conditioning on x_t rather than using a full model for the data generation process of all the variables involved. In that case, x_t is called *weakly exogenous* for $\boldsymbol{\gamma}$. This and other types of exogeneity have been formalized by Engle, Hendry & Richard (1983). They call x_t *strongly exogenous* if we can condition on this set of variables for forecasting purposes without loosing forecast precision and they classify x_t as *super-exogenous* if policy analysis can be made conditional on these variables (see also Geweke (1982), Hendry (1995, Chapter 5), Ericsson (1994) for more discussion of exogeneity).

A simple technical definition is to call x_t exogenous if $x_t, x_{t-1}, \dots, x_{t-s}$ are independent of the error term w_t. Moreover, x_t is sometimes called *strictly exogenous* if all its leads and lags are independent of all leads and lags of the error process w_t, that is, if x_t and w_t are independent processes. Such assumptions simplify derivations of properties of estimators and are therefore convenient. They may be unnecessarily restrictive, however, for some purposes. In the following, we will implicitly make the assumption that x_t and w_t are independent processes for convenience, although most results can be obtained under less restrictive conditions.

For much of the present discussion, a formal definition of the types of variables involved is not necessary. It suffices to have a partitioning into two groups of variables. The reader should, however, have some intuition of which variables are contained in y_t and which ones are included in x_t. As mentioned previously, roughly speaking, y_t contains the observable *outputs* of the system, that is, the observable variables that are determined by the system. In contrast, the x_t variables may be regarded as *observable input variables* which are determined outside the system. In this setting, the error variables w_t may be viewed as *unobservable inputs* to the system. As we have seen, nonstochastic components may be absorbed into the set of x_t variables. All or some of the components of x_t may be under full or partial control of the government or a decision or policy maker. In a control context, such variables are often referred to as *instruments* or *instrument variables* (see Section 10.7). Sometimes the lagged endogenous variables together with the exogenous variables of a system are called *predetermined variables*. If x_t contains just a constant and $s = 0$, the model (10.1.1) reduces to a VAR model, provided w_t is white noise.

For illustrative purposes, consider the following example system relating investment (x_{1t}), income (y_{1t}), and consumption (y_{2t}) variables:

$$\begin{aligned} y_{1t} &= \nu_1^* + \alpha_{11,1}^* y_{1,t-1} + \alpha_{12,1}^* y_{2,t-1} + \beta_{12,1}^* x_{1,t-1} + w_{1t}, \\ y_{2t} &= \nu_2^* + \alpha_{22,1}^* y_{2,t-1} + \mathsf{a}_{21,0} y_{1t} + \alpha_{21,1}^* y_{1,t-1} + w_{2t}. \end{aligned} \tag{10.2.1}$$

This model is similar to those obtained for West German data in Chapter 5. An important difference is that current income appears in the consumption equation and there is no equation for investment. Thus, only income and consumption are determined within the system whereas investment is not. The fact that investment is, of course, determined within the economic system as a whole does not necessarily mean that we have to specify its generation mechanism if our main interest is with the generation mechanism of income and consumption. In terms of the representation (10.1.1), the example system can be written as

$$\begin{aligned} \begin{bmatrix} 1 & 0 \\ -\mathsf{a}_{21,0} & 1 \end{bmatrix} \begin{bmatrix} y_{1t} \\ y_{2t} \end{bmatrix} = & \begin{bmatrix} \alpha_{11,1}^* & \alpha_{12,1}^* \\ \alpha_{21,1}^* & \alpha_{22,1}^* \end{bmatrix} \begin{bmatrix} y_{1,t-1} \\ y_{2,t-1} \end{bmatrix} \\ & + \begin{bmatrix} \nu_1^* & \beta_{12,1}^* \\ \nu_2^* & 0 \end{bmatrix} \begin{bmatrix} 1 \\ x_{1,t-1} \end{bmatrix} + \begin{bmatrix} w_{1t} \\ w_{2t} \end{bmatrix}. \end{aligned} \tag{10.2.2}$$

Thus, $y_t = (y_{1t}, y_{2t})'$ and $x_t = (1, x_{1t})'$ are both two-dimensional. The predetermined variables are y_{t-1} and x_{t-1}.

In dynamic SEMs there are sometimes identities or exact relations between some variables. For instance, the same figures may be used for supply and demand of a product. In that case, an identity equating supply and demand may appear as a separate equation of a system. So far we have not excluded this possibility. However, in later sections the covariance matrix of w_t will be assumed to be nonsingular which excludes identities. Then we assume without further notice that they have been eliminated by substitution. For instance, the demand variable may be substituted for the supply variable in all instances where it appears in the system.

10.2.2 Structural Form, Reduced Form, Final Form

The representation (10.1.1) is called the *structural form* of the model if it represents the instantaneous effects of the endogenous variables properly. The instantaneous effects are reflected in the elements of A. The idea is that the instantaneous causal links are derived from theoretical considerations and are used to place restrictions on A. Of course, multiplication of (10.1.1) with any other nonsingular $(K \times K)$ matrix results in an equivalent representation of the process generating y_t. Such a representation is not called a structural form, however, unless it reflects the actual relations of interest.

The *reduced form* of the system is obtained by premultiplying (10.1.1) with A^{-1} which gives

$$y_t = A_1 y_{t-1} + \cdots + A_p y_{t-p} + B_0 x_t + \cdots + B_s x_{t-s} + u_t, \qquad (10.2.3)$$

where $A_i := \mathsf{A}^{-1} A_i^*$ $(i = 1, \ldots, p)$, $B_j := \mathsf{A}^{-1} B_j^*$ $(j = 0, 1, \ldots, s)$, and $u_t := \mathsf{A}^{-1} w_t$. We always assume without notice that the inverse of A exists. In Sections 10.5–10.7, we will see that the reduced form is useful for forecasting, multiplier analysis, and control purposes.

For the example model given in (10.2.2), we have

$$\mathsf{A}^{-1} = \begin{bmatrix} 1 & 0 \\ \mathsf{a}_{21,0} & 1 \end{bmatrix}$$

and, hence, the reduced form is

$$\begin{bmatrix} y_{1t} \\ y_{2t} \end{bmatrix} = A_1 \begin{bmatrix} y_{1,t-1} \\ y_{2,t-1} \end{bmatrix} + B_1 \begin{bmatrix} 1 \\ x_{1,t-1} \end{bmatrix} + \begin{bmatrix} u_{1t} \\ u_{2t} \end{bmatrix}, \qquad (10.2.4)$$

where

$$A_1 = \begin{bmatrix} \alpha_{11,1} & \alpha_{12,1} \\ \alpha_{21,1} & \alpha_{22,1} \end{bmatrix} = \begin{bmatrix} \alpha_{11,1}^* & \alpha_{12,1}^* \\ \mathsf{a}_{21,0}\alpha_{11,1}^* + \alpha_{21,1}^* & \mathsf{a}_{21,0}\alpha_{12,1}^* + \alpha_{22,1}^* \end{bmatrix}, \qquad (10.2.5)$$

$$B_1 = \begin{bmatrix} \beta_{11,1} & \beta_{12,1} \\ \beta_{21,1} & \beta_{22,1} \end{bmatrix} = \begin{bmatrix} \nu_1^* & \beta_{12,1}^* \\ \mathsf{a}_{21,0}\nu_1^* + \nu_2^* & \mathsf{a}_{21,0}\beta_{12,1}^* \end{bmatrix}, \tag{10.2.6}$$

and

$$\begin{bmatrix} u_{1t} \\ u_{2t} \end{bmatrix} = \begin{bmatrix} w_{1t} \\ \mathsf{a}_{21,0}w_{1t} + w_{2t} \end{bmatrix}.$$

It is important to note that the reduced form parameters are in general nonlinear functions of the structural form parameters.

In lag operator notation, the reduced form (10.2.3) can be written as

$$A(L)y_t = B(L)x_t + u_t, \tag{10.2.7}$$

where

$$A(L) := I_K - A_1L - \cdots - A_pL^p$$

and

$$B(L) := B_0 + B_1L + \cdots + B_sL^s.$$

If the effect of a change in an exogenous variable on the endogenous variables is of interest, it is useful to solve the system (10.2.7) for the endogenous variables by multiplying with $A(L)^{-1}$. The resulting representation,

$$y_t = D(L)x_t + A(L)^{-1}u_t, \tag{10.2.8}$$

where $D(L) := A(L)^{-1}B(L)$, is sometimes called the *final form* of the system. Of course, using $A(L)^{-1}$ requires invertibility of $A(L)$ which is guaranteed if

$$\det A(z) \neq 0 \quad \text{for } |z| \leq 1. \tag{10.2.9}$$

If y_t contains just one variable, $A(L)$ is a scalar operator and the form (10.2.8) is often called a *distributed lag model* in the econometrics literature because it describes how lagged effects of changes in x_t are distributed over time. Because the lag distribution for each exogenous variable can be written as a ratio of two finite order polynomials in the lag operator $(A(L)^{-1}B(L))$, the model is referred to as a *rational distributed lag model*. In the time series literature, the label *rational transfer function model* is often attached to (10.2.8) in both the scalar and the vector case. The operator $D(L)$ represents the *transfer function* transferring the observable inputs into the outputs of the system.

For the example model with reduced form (10.2.4), we get a final form

$$\begin{aligned} \begin{bmatrix} y_{1t} \\ y_{2t} \end{bmatrix} &= (I_2 - A_1L)^{-1}B_1L\begin{bmatrix} 1 \\ x_{1t} \end{bmatrix} + (I_2 - A_1L)^{-1}\begin{bmatrix} u_{1t} \\ u_{2t} \end{bmatrix} \\ &= \left(\sum_{i=1}^{\infty} A_1^{i-1}B_1L^i\right)\begin{bmatrix} 1 \\ x_{1t} \end{bmatrix} + \left(\sum_{i=0}^{\infty} A_1^iL^i\right)\begin{bmatrix} u_{1t} \\ u_{2t} \end{bmatrix}. \end{aligned} \tag{10.2.10}$$

Note that $B_0 = 0$ and thus, $D_0 = 0$ and $D_i = A_1^{i-1} B_1$ for $i = 1, 2, \ldots$.

The coefficient matrices $D_i = (d_{kj,i})$ of the transfer function operator

$$D(L) = \sum_{i=0}^{\infty} D_i L^i$$

contain the effects that changes in the exogenous variables have on the endogenous variables. Everything else held constant, a unit change in the j-th exogenous variable in period t induces a marginal change of $d_{kj,i}$ units in the k-th endogenous variable in period $t + i$. The elements of the D_i matrices are therefore called *dynamic multipliers*. The accumulated effects contained in $\sum_{i=0}^{n} D_i$ are the n-th *interim multipliers* and the elements of $\sum_{i=0}^{\infty} D_i$ are the *long-run effects* or *total multipliers*. We will return to multiplier analysis in Section 10.6.

As in the example, the transfer function operator $D(L)$ has infinite order in general. A finite order representation of the system is obtained by noting that $A(L)^{-1} = A(L)^{adj}/|A(L)|$, where $A(L)^{adj}$ denotes, as usual, the adjoint of $A(L)$. Thus, multiplying the reduced form by $A(L)^{adj}$ gives

$$|A(L)| y_t = A(L)^{adj} B(L) x_t + A(L)^{adj} u_t \tag{10.2.11}$$

which involves finite order operators only. In the econometrics literature these equations are sometimes called *final equations*. Because $|A(L)|$ is a scalar operator, each equation contains only one of the endogenous variables.

Assuming that the unmodelled variables x_t are driven by a VAR(q) process, say

$$x_t = C_1 x_{t-1} + \cdots + C_q x_{t-q} + v_t,$$

where $q \leq p$ and v_t is white noise, then the joint generation process of x_t and y_t is

$$\begin{bmatrix} I_K & -B_0 \\ 0 & I_M \end{bmatrix} \begin{bmatrix} y_t \\ x_t \end{bmatrix} = \begin{bmatrix} A_1 & B_1 \\ 0 & C_1 \end{bmatrix} \begin{bmatrix} y_{t-1} \\ x_{t-1} \end{bmatrix} + \cdots + \begin{bmatrix} A_p & B_p \\ 0 & C_p \end{bmatrix} \begin{bmatrix} y_{t-p} \\ x_{t-p} \end{bmatrix} + \begin{bmatrix} u_t \\ v_t \end{bmatrix},$$

where it is assumed without loss of generality that $s, q \leq p$, $B_i := 0$ for $i > s$ and $C_j := 0$ for $j > q$. If u_t is also white noise, premultiplying by

$$\begin{bmatrix} I_K & -B_0 \\ 0 & I_M \end{bmatrix}^{-1} = \begin{bmatrix} I_K & B_0 \\ 0 & I_M \end{bmatrix}$$

shows that the joint generation process of y_t and x_t is a VAR(p).

10.2.3 Models with Rational Expectations

Sometimes the endogenous variables are assumed to depend not only on other endogenous and exogenous variables but also on expectations on endogenous variables. If only expectations formed in the previous period for the present period are of importance, one could simply add another term involving the expectations variables to the structural form (10.1.1). Denoting the expectations variables by y_t^e may then result in a reduced form

$$y_t = A_1 y_{t-1} + \cdots + A_p y_{t-p} + F y_t^e + B_0 x_t + \cdots + B_s x_{t-s} + u_t \quad (10.2.12)$$

or

$$A(L) y_t = F y_t^e + B(L) x_t + u_t, \quad (10.2.13)$$

where F is a $(K \times K)$ matrix of parameters and $A(L)$ and $B(L)$ are the matrix polynomials in the lag operator from (10.2.7).

Following Muth (1961), the expectations y_t^e formed in period $t-1$ are called *rational* if they are the best possible predictions, given the information in period $t-1$. In other words, y_t^e is the conditional expectation $E_{t-1}(y_t)$, given all information available in period $t-1$. In forming the predictions or expectations, not only the past values of the endogenous and unmodelled variables are assumed to be known but also the model (10.2.12) and the generation process of the unmodelled variables. It is easy to see that, if the unmodelled variables are generated by a VAR process, the expectations variables can be eliminated from (10.2.12)/(10.2.13). The resulting reduced form is of VARX type. To show this result, suppose that u_t is independent white noise and, as before, denote by E_t the conditional expectation, given all information available in period t. Applying E_{t-1} to (10.2.12) then gives

$$\begin{aligned} y_t^e &= E_{t-1}(y_t) \\ &= A_1 y_{t-1} + \cdots + A_p y_{t-p} \\ &\quad + F y_t^e + B_0 E_{t-1}(x_t) + B_1 x_{t-1} + \cdots + B_s x_{t-s} \end{aligned} \quad (10.2.14)$$

or

$$y_t^e = (A(L) - I_K) y_t + F y_t^e + B_0 E_{t-1}(x_t) + (B(L) - B_0) x_t. \quad (10.2.15)$$

Assuming that $I_K - F$ is invertible, this system can be solved for y_t^e:

$$y_t^e = (I_K - F)^{-1} [(A(L) - I_K) y_t + B_0 E_{t-1}(x_t) + (B(L) - B_0) x_t]. \quad (10.2.16)$$

If x_t is generated by a VAR(q) process, say

$$x_t = C_1 x_{t-1} + \cdots + C_q x_{t-q} + v_t,$$

where v_t is independent white noise, then

$$E_{t-1}(x_t) = C_1 x_{t-1} + \cdots + C_q x_{t-q}.$$

Substituting this expression in (10.2.16) shows that y_t^e depends on lagged y_t and x_t only. Thus, substituting for y_t^e in (10.2.12) or (10.2.13), we get a standard VARX form of the model.

Thus, in theory, when the true coefficient matrices are known, we can simply eliminate the term involving expectations variables and work with a standard reduced form without an expectations term. It should be clear, however, that substituting the right-hand side of (10.2.16) for y_t^e in (10.2.12) implies nonlinear restrictions on the coefficient matrices of the reduced form without expectations terms. Taking into account such restrictions may increase the efficiency of parameter estimators. The same is true, of course, for the structural form. Therefore, it is important in practice whether or not the actual relationship between the variables is partly determined by agents' expectations.

For expository purposes we have just treated a very special case where only expectations formed in period $t-1$ for period t enter the model. Extensions can be treated in a similar way. For instance, past expectations for more than one period ahead or expectations formed in various previous periods may be of importance. If x_t is generated by a VAR(q) process, they can be eliminated like in the special case considered in the foregoing.

A complication of the basic model that makes life a bit more difficult is the inclusion of future expectations. It is quite realistic to suppose that, for instance, the expected future price of a commodity may determine the supply in the present period. For example, if bond prices are expected to fall during the next period, an investor may decide to sell now. If future expectations enter the model, the solution for the endogenous variables will in general not be unique. In other words, the process that generates the endogenous variables may not be uniquely determined by the model, even if the generation process of the exogenous variables is uniquely specified. Further extensive discussions of rational expectations models can be found in volumes by Lucas & Sargent (1981) and Pesaran (1987).

10.2.4 Cointegrated Variables

Many of the results discussed so far in this section hold for systems of stationary or integrated variables. More precisely, whenever the VAR operator $A(L)$ is not required to be invertible, integrated variables may be present as endogenous as well as unmodelled variables. If there are cointegrated variables, it may be preferable, however, to separate the short- and long-run dynamics as in a VECM. Assuming that there are r cointegration relations among the endogenous variables and they are not cointegrated with the unmodelled variables, the corresponding form of the model is

$$\begin{aligned} \mathsf{A}\Delta y_t &= \alpha^* \beta' y_{t-1} + \mathbf{\Gamma}_1^* \Delta y_{t-1} + \cdots + \mathbf{\Gamma}_{p-1}^* \Delta y_{t-p+1} \\ &\quad + B_0^* x_t + B_1^* x_{t-1} + \cdots + B_s^* x_{t-s} + w_t, \end{aligned} \quad (10.2.17)$$

where A is a $(K \times K)$ matrix of instantaneous effects, as before, α^* is a $(K \times r)$ matrix of structural loading coefficients, β is the $(K \times r)$ cointegration matrix, $\mathbf{\Gamma}_j^*$ $(j = 1, \dots, p-1)$ is a $(K \times K)$ matrix of structural short-run coefficients, and all other symbols are defined as in (10.1.1). In many respects, this model can be dealt with in essentially the same way as the VECMs considered in Part II of this volume.

It is also possible, however, that there is cointegration between endogenous and unmodelled variables. In that case, a suitable form of the model is

$$\begin{aligned} \mathsf{A}\Delta y_t &= \alpha^*\beta^{+\prime}\begin{bmatrix} y_{t-1} \\ x_{t-1} \end{bmatrix} + \mathbf{\Gamma}_1^*\Delta y_{t-1} + \cdots + \mathbf{\Gamma}_{p-1}^*\Delta y_{t-p+1} \\ &\quad + \mathbf{\Upsilon}_0^*\Delta x_t + \mathbf{\Upsilon}_1^*\Delta x_{t-1} + \cdots + \mathbf{\Upsilon}_{s-1}^*\Delta x_{t-s+1} + w_t, \end{aligned} \tag{10.2.18}$$

where now the unmodelled variables appear in levels form in the error correction term only and otherwise enter in differenced form with suitable coefficient matrices $\mathbf{\Upsilon}_j^*$ $(j = 0, 1, \dots, s-1)$. It is easy to see that such a model form can be obtained if the joint generation process of y_t and x_t has a (reduced form) VECM representation

$$\begin{aligned} \begin{bmatrix} \Delta y_t \\ \Delta x_t \end{bmatrix} &= \begin{bmatrix} \alpha \\ \alpha_x \end{bmatrix}\beta^{+\prime}\begin{bmatrix} y_{t-1} \\ x_{t-1} \end{bmatrix} + \begin{bmatrix} \mathbf{\Gamma}_1 & \mathbf{\Upsilon}_1 \\ 0 & \mathbf{\Gamma}_1^x \end{bmatrix}\begin{bmatrix} \Delta y_{t-1} \\ \Delta x_{t-1} \end{bmatrix} + \cdots \\ &\quad + \begin{bmatrix} \mathbf{\Gamma}_{p-1} & \mathbf{\Upsilon}_{p-1} \\ 0 & \mathbf{\Gamma}_{p-1}^x \end{bmatrix}\begin{bmatrix} \Delta y_{t-p+1} \\ \Delta x_{t-p+1} \end{bmatrix} + \begin{bmatrix} u_t \\ v_t \end{bmatrix}, \end{aligned} \tag{10.2.19}$$

where $p \geq s$ is assumed without loss of generality and all symbols have obvious definitions. Premultiplying this model form with

$$\begin{bmatrix} \mathsf{A} & -\mathbf{\Upsilon}_0^* \\ 0 & I_M \end{bmatrix}$$

gives a model where the first K equations are just the structural form (10.2.18). Notice, however, that the y_t may enter the x_t equations in (10.2.19) via the cointegration relations if $\alpha_x \neq 0$. It turns out that x_t is weakly exogenous for β^+, if $\alpha_x = 0$. Thus, if the cointegration relations are of primary interest, considering the partial model for Δy_t is justified if $\alpha_x = 0$.

Both models (10.2.17) and (10.2.18) can be rewritten in levels form. The result is then a structural form as in (10.1.1). Moreover, the structural forms can be converted into reduced form by premultiplying with A^{-1}.

10.3 Estimation

Parameter estimation in the presence of unmodelled variables will be discussed separately for stationary and cointegrated variables. We begin with the stationary case.

10.3.1 Stationary Variables

Suppose $(y_t', x_t')'$ is generated by a stationary process and we wish to estimate the parameters of the reduced form (10.2.3) which can be written as

$$y_t = AY_{t-1} + BX_{t-1} + B_0 x_t + u_t, \tag{10.3.1}$$

where $A := [A_1, \ldots, A_p]$, $B := [B_1, \ldots, B_s]$,

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

Here u_t is assumed to be standard white noise with *nonsingular* covariance matrix Σ_u. Moreover, we allow for parameter restrictions and assume that a matrix R and a vector $\boldsymbol{\gamma}$ exist such that

$$\boldsymbol{\beta} := \operatorname{vec}[A, B, B_0] = R\boldsymbol{\gamma}. \tag{10.3.2}$$

With these assumptions, estimation of $\boldsymbol{\beta}$ and, hence, of A, B, and B_0 is straightforward.

For a sample of size T, the system can be written compactly as

$$Y = [A, B, B_0]Z + U, \tag{10.3.3}$$

where

$$Y := [y_1, \ldots, y_T], \quad Z := \begin{bmatrix} Y_0, \ldots, Y_{T-1} \\ X_0, \ldots, X_{T-1} \\ x_1, \ldots, x_T \end{bmatrix} \quad \text{and} \quad U := [u_1, \ldots, u_T].$$

Vectorizing gives

$$\mathbf{y} = (Z' \otimes I_K)R\boldsymbol{\gamma} + \mathbf{u},$$

where $\mathbf{y} := \operatorname{vec}(Y)$ and $\mathbf{u} := \operatorname{vec}(U)$. From Chapter 5, the GLS estimator is known to be

$$\widehat{\boldsymbol{\gamma}} = [R'(ZZ' \otimes \Sigma_u^{-1})R]^{-1}R'(Z \otimes \Sigma_u^{-1})\mathbf{y}. \tag{10.3.4}$$

This estimator is not operational because in practice Σ_u is unknown. However, as in Section 5.2.2, Σ_u may be estimated from the LS estimator

$$\breve{\boldsymbol{\gamma}} = [R'(ZZ' \otimes I_K)R]^{-1}R'(Z \otimes I_K)\mathbf{y}$$

which gives residuals $\breve{\mathbf{u}} = \mathbf{y} - (Z' \otimes I_K)R\breve{\boldsymbol{\gamma}}$ and an estimator

$$\breve{\Sigma}_u = \breve{U}\breve{U}'/T \tag{10.3.5}$$

of Σ_u, where $\breve{U}$ is such that $\text{vec}(\breve{U}) = \breve{\mathbf{u}}$. Using this estimator of the white noise covariance matrix results in the EGLS estimator

$$\widehat{\widehat{\boldsymbol{\gamma}}} = [R'(ZZ' \otimes \breve{\Sigma}_u^{-1})R]^{-1}R'(Z \otimes \breve{\Sigma}_u^{-1})\mathbf{y}. \tag{10.3.6}$$

Under standard assumptions, this estimator is consistent and asymptotically normal,

$$\sqrt{T}(\widehat{\widehat{\boldsymbol{\gamma}}} - \boldsymbol{\gamma}) \xrightarrow{d} \mathcal{N}(0, \Sigma_{\widehat{\widehat{\boldsymbol{\gamma}}}}), \tag{10.3.7}$$

where

$$\Sigma_{\widehat{\widehat{\boldsymbol{\gamma}}}} = (R'[\text{plim}(T^{-1}ZZ') \otimes \Sigma_u^{-1}]R)^{-1}. \tag{10.3.8}$$

One condition for this result to hold is, of course, that both plim $T^{-1}ZZ'$ and the inverse of the matrix in (10.3.8) exist. Further assumptions are required to guarantee the asymptotic normal distribution of the EGLS estimator. The assumptions may include the following ones: (i) u_t is standard white noise, (ii) the VAR part is stable, that is,

$$|A(z)| = |I_K - A_1 z - \cdots - A_p z^p| \neq 0 \quad \text{for } |z| \leq 1,$$

and (iii) x_t is generated by a stationary, stable VAR process which is independent of the white noise process u_t. A precise statement of more general conditions and a proof are given, e.g., by Hannan & Deistler (1988). The latter part of our set of assumptions requires that all the exogenous variables are stochastic. It can be modified so as to include nonstochastic variables as well. In that case, the plim in (10.3.8) reduces to a nonstochastic limit in some or all components (see, e.g., Anderson (1971, Chapter 5), Harvey (1981)).

An estimator for $\boldsymbol{\beta} = R\boldsymbol{\gamma}$ is obtained as $\widehat{\widehat{\boldsymbol{\beta}}} = R\widehat{\widehat{\boldsymbol{\gamma}}}$. If (10.3.7) holds, this estimator also has an asymptotic normal distribution,

$$\sqrt{T}(\widehat{\widehat{\boldsymbol{\beta}}} - \boldsymbol{\beta}) \xrightarrow{d} \mathcal{N}(0, \Sigma_{\widehat{\widehat{\boldsymbol{\beta}}}} = R\Sigma_{\widehat{\widehat{\boldsymbol{\gamma}}}}R'), \tag{10.3.9}$$

Moreover, under general conditions, the corresponding estimator $\widehat{\widehat{\Sigma}}_u$ of the white noise covariance matrix is asymptotically independent of $\widehat{\widehat{\boldsymbol{\beta}}}$ and has the same asymptotic distribution as the estimator UU'/T based on the unobserved true residuals. For instance, for a Gaussian process,

$$\sqrt{T}\,\text{vech}(\widehat{\widehat{\Sigma}}_u - \Sigma_u) \xrightarrow{d} \mathcal{N}(0, 2\mathbf{D}_K^+(\Sigma_u \otimes \Sigma_u)\mathbf{D}_K^{+\prime}), \tag{10.3.10}$$

where $\mathbf{D}_K^+ = (\mathbf{D}_K'\mathbf{D}_K)^{-1}\mathbf{D}_K'$ is the Moore-Penrose inverse of the $(K^2 \times \frac{1}{2}K(K+1))$ duplication matrix $\mathbf{D}_K$.

In discussing direct reduced form estimation with white noise errors, we have treated a particularly simple case. The following complications are possible.

(1) Usually there will be restrictions on the structural form coefficients A, A_i^*, $i = 1, \dots, p$, and B_j^*, $j = 0, \dots, s$. Such restrictions may imply nonlinear constraints on the reduced form coefficients which are not covered by the above approach. Rational expectations assumptions may be another source of nonlinear restrictions on the reduced form parameters. Theoretically, it is not difficult to handle nonlinear restrictions on the reduced form parameters. In practice, numerical problems may arise in a multivariate LS or GLS estimation with nonlinear restrictions.
(2) Interest may focus on the structural rather than the reduced form. Estimation of the structural form has been discussed extensively in the econometrics literature. For recent surveys and many further references see Judge et al. (1985), Hausman (1983), or textbooks such as Hayashi (2000). A major complication in estimating the structural form of a SEM such as (10.1.1) results from its possible nonuniqueness. Note that we have not assumed a triangular A matrix or a diagonal covariance matrix of w_t. Premultiplication of (10.1.1) by any nonsingular matrix results in an equivalent representation of the process. Thus, for proper estimation there must be restrictions on the structural form coefficients that guarantee uniqueness or identification of the structural form coefficients.
(3) So far we have just discussed models which are linear in the variables. In practice, there may be nonlinear relations between the variables. Estimation of nonlinear dynamic models where the endogenous as well as the unmodelled conditioning variables may enter in a nonlinear way are, for instance, discussed by Bierens (1981), Gallant (1987), and Gallant & White (1988).

In the next section, we will consider models with integrated and cointegrated variables.

10.3.2 Estimation of Models with $I(1)$ Variables

If there are integrated and cointegrated variables in the model and a reduced form VECM corresponding to the structural form (10.2.18),

$$\begin{aligned} \Delta y_t &= \alpha\beta^{+\prime} \begin{bmatrix} y_{t-1} \\ x_{t-1} \end{bmatrix} + \mathbf{\Gamma}_1 \Delta y_{t-1} + \cdots + \mathbf{\Gamma}_{p-1} \Delta y_{t-p+1} \\ &\quad + \mathbf{\Upsilon}_0 \Delta x_t + \mathbf{\Upsilon}_1 \Delta x_{t-1} + \cdots + \mathbf{\Upsilon}_{s-1} \Delta x_{t-s+1} + u_t, \end{aligned} \tag{10.3.11}$$

is set up, estimation can in principle proceed as in Section 7.2. Assuming that a sample of size T and all required presample values are available and defining

$$\Delta Y := [\Delta y_1, \dots, \Delta y_T],$$

$$Y_{-1}^+ := [y_0^+, \dots, y_{T-1}^+], \quad \text{with} \quad y_{t-1}^+ := \begin{bmatrix} y_{t-1} \\ x_{t-1} \end{bmatrix},$$

$$\Delta X^+ := [\Delta X_0^+, \dots, \Delta X_{T-1}^+] \quad \text{with} \quad \Delta X_{t-1}^+ := \begin{bmatrix} \Delta y_{t-1} \\ \vdots \\ \Delta y_{t-p+1} \\ \Delta x_t \\ \Delta x_{t-1} \\ \vdots \\ \Delta x_{t-s+1} \end{bmatrix},$$

and

$$U := [u_1, \dots, u_T],$$

we get

$$\Delta Y = \alpha\beta^{+\prime} Y_{-1}^+ + \mathbf{\Gamma}^+ \Delta X^+ + U, \tag{10.3.12}$$

where

$$\mathbf{\Gamma}^+ := [\mathbf{\Gamma}_1 : \dots : \mathbf{\Gamma}_{p-1} : \mathbf{\Upsilon}_0 : \mathbf{\Upsilon}_1 : \dots : \mathbf{\Upsilon}_{s-1}].$$

Thus, we have precisely the same model form as in Section 7.2 (see, e.g., (7.2.3)) and, in principle, all the estimators of that section are available. Notice, however, that now β^+ is a $((K + M) \times r)$ matrix whereas α is still $(K \times r)$. Because the error correction term now involves all the cointegration relations between the endogenous and unmodelled variables, it is possible that $r > K$. In that case, it is easy to see that most of the estimators of Section 7.2 are not available. Thus, we have to assume that $r \leq K$. In fact, if $r = K$, the matrix $\mathbf{\Pi}^+ := \alpha\beta^{+\prime}$ is of full row rank under our usual assumption that $\text{rk}(\alpha) = \text{rk}(\beta^+) = r$. Therefore, if $K = r$, we do not even need reduced rank regression but can simply estimate the matrix $\mathbf{\Pi}^+ = \alpha\beta^{+\prime}$ by applying multivariate LS to (10.3.12). An estimator of β^+ can then be obtained by normalizing the cointegration matrix as in Section 7.2 such that

$$\beta^+ = \begin{bmatrix} I_K \\ \beta_{(M)}^+ \end{bmatrix} \tag{10.3.13}$$

and, using

$$\widehat{\beta}^{+\prime} = (\widehat{\mathbf{\Pi}}_{(1)}^+)^{-1} \widehat{\mathbf{\Pi}}^+,$$

where $\widehat{\mathbf{\Pi}}_{(1)}^+$ is the $(K \times K)$ submatrix consisting of the first K columns of the LS estimator $\widehat{\mathbf{\Pi}}^+$ of $\mathbf{\Pi}^+$.

If $r < K$, there is nothing special here relative to the procedures discussed in Section 7.2. Reduced rank ML estimation, as discussed in Section 7.2.3, is available just as the EGLS estimator of the cointegration parameters of Section 7.2.2 and the two-stage estimator described in Section 7.2.5. Moreover, the two-stage procedure can also be used to estimate models with parameter restrictions on α and $\mathbf{\Gamma}^+$, as in Section 7.3.2. In fact, a similar procedure can even be used for the estimation of structural form models of the type (10.2.18).

In this context, it is, of course, of interest to know the properties of the resulting estimators. They are available under suitable assumptions for the model and the variables (see, e.g., Johansen (1992) or Davidson (2000, Section 16.5)). Under general assumptions, the estimator of the cointegration matrix continues to be superconsistent, that is,

$$T(\widehat{\beta}^+ - \beta^+) = O_p(1),$$

if all variables are at most $I(1)$ and β^+ is identified. If the cointegration relations do not enter the generation process of x_t, that is, $\alpha_x = 0$ in (10.2.19), x_t is weakly exogenous for β^+ and the ML and EGLS estimators of β^+ have mixed normal distributions similar to those discussed in Section 7.2. Therefore standard inference is possible, as discussed in that section. The estimators of the α and $\mathbf{\Gamma}^+$ parameters have again standard properties which are the same as in the case where the β^+ matrix is known.

10.4 Remarks on Model Specification and Model Checking

The basic principles of model specification and checking the model adequacy have been discussed in some detail in previous chapters. We will therefore make just a few remarks here. With respect to the specification there is, however, a major difference between the models considered previously and the dynamic SEMs of this chapter. While in a reduced form VAR analysis usually relatively little prior knowledge from economic or other subject matter theory is used, such theories may well be the major building block in specifying SEMs. In that case, model checking becomes of central importance in investigating the validity of the theory. Quite often, theories are not available that specify the data generation process completely. For instance, the lag lengths of the endogenous and/or exogenous variables may have to be specified with statistical tools. Also, some researchers may not be prepared to rely on the available theories and therefore prefer to substitute statistical investigations for uncertain prior knowledge. Statistical specification strategies for general dynamic SEMs were, for instance, proposed and discussed by Hannan & Kavalieris (1984), Hannan & Deistler (1988), and Poskitt (1992). These strategies are based on model selection criteria of the type considered in previous chapters. An extensive literature exists on the specification of special models. For instance, distributed lag models are discussed at length in the econometrics literature (for some references see Judge et al. (1985, Chapters 9 and 10)). Specification proposals for transfer function models with one dependent variable y_t go back to the pioneering work of Box & Jenkins (1976). Other suggestions have been made by Haugh & Box (1977), Young, Jakeman & McMurtrie (1980), Liu & Hanssens (1982), Tsay (1985), and Poskitt (1989) to name just a few.

If some of the variables are integrated, one may also want to investigate the number of cointegration relations with statistical tests. From the discussion in Section 10.3.2, it is clear that rank tests can be used for that purpose, as in Section 8.2. These tests may now be based either on a VECM for the full joint generation process of y_t and x_t or on a partial model with some unmodelled variables. The latter approach may be preferable if a large number of variables is involved. Johansen's LR tests for the cointegrating rank may be unreliable in that situation because of size distortions and lack of power. Therefore, testing for the cointegrating rank in a partial model may be advantageous. The asymptotic distributions of the relevant LR test statistics in this case depend on the conditioning variables, however. This result is not surprising, of course, because the conditioning variables can in fact be deterministic terms and we have seen in Section 8.2 that such terms have an impact on the asymptotic properties of the LR tests. The relevant tests for conditional models were derived by Harbo, Johansen, Nielsen & Rahbek (1998) and critical values were given in MacKinnon, Haug & Michelis (1999).

In checking the model adequacy one may want to test various restrictions. These may range from constraints suggested by some kind of theory such as the rational expectations hypothesis, to tests of the significance of extra lags. The three testing principles discussed previously, namely the LR, LM, and Wald principles (see Appendix C.7) can be used in the present context. Their asymptotic properties follow in the usual way from properties of the estimators and the model.

A residual analysis is another tool which is available in the present case. Plots of residuals may help to identify unusual values or patterns that suggest model deficiencies. Plots of residual autocorrelations may aid in checking the white noise assumption. Also a portmanteau test for overall residual autocorrelation may be developed for dynamic models with exogenous variables; see Poskitt & Tremayne (1981) for a discussion of this issue and further references.

10.5 Forecasting

10.5.1 Unconditional and Conditional Forecasts

If the future paths of the unmodelled variables are unknown to the forecaster, then forecasts of these variables are needed in order to predict the future values of the endogenous variables on the basis of a dynamic SEM. For simplicity, suppose that the exogenous variables are generated by a zero mean VAR(q) process as in Section 10.2.3,

$$x_t = C_1 x_{t-1} + \cdots + C_q x_{t-q} + v_t. \tag{10.5.1}$$

Now this process can be used to produce optimal forecasts $x_t(h)$ of x_t in the usual way. If the endogenous variables are generated by the reduced form model (10.2.3) with u_t being independent white noise which is also independent of the x_t process, the optimal h-step forecast of y_{t+h} at origin t is

$$y_t(h) = A_1 y_t(h-1) + \cdots + A_p y_t(h-p) + B_0 x_t(h) + \cdots + B_s x_t(h-s), \quad (10.5.2)$$

where $y_t(j) := y_{t+j}$ and $x_t(j) := x_{t+j}$ for $j \leq 0$. This formula can be used for recursively determining forecasts for $h = 1, 2, \ldots$.

An alternative way for getting these forecasts is obtained by writing the generation processes of the exogenous variables in one overall model together with the reduced form SEM:

$$\begin{aligned} \begin{bmatrix} I_K & -B_0 \\ 0 & I_M \end{bmatrix} \begin{bmatrix} y_t \\ x_t \end{bmatrix} &= \begin{bmatrix} A_1 & B_1 \\ 0 & C_1 \end{bmatrix} \begin{bmatrix} y_{t-1} \\ x_{t-1} \end{bmatrix} + \cdots \\ &\quad + \begin{bmatrix} A_p & B_p \\ 0 & C_p \end{bmatrix} \begin{bmatrix} y_{t-p} \\ x_{t-p} \end{bmatrix} + \begin{bmatrix} u_t \\ v_t \end{bmatrix}, \end{aligned} \quad (10.5.3)$$

where we assume without loss of generality that $p \geq \max(s, q)$ and set $B_i = 0$ for $i > s$ and $C_j = 0$ for $j > q$. As in Section 10.2.2, premultiplying by

$$\begin{bmatrix} I_K & -B_0 \\ 0 & I_M \end{bmatrix}^{-1} = \begin{bmatrix} I_K & B_0 \\ 0 & I_M \end{bmatrix}$$

gives a standard reduced form VAR(p) model. It is easy to see that the optimal forecasts for y_t and x_t from that model are exactly the same as those obtained by getting forecasts for x_t from (10.5.1) first and using them in the prediction formula for y_t given in (10.5.2) (see Problem 10.5). Thus, under the present assumptions, the discussion of forecasting VAR(p) processes applies. It will not be repeated here. Also, it is not difficult to extend these ideas to sets of unmodelled variables with nonstochastic components such as intercept terms or seasonal dummies.

We will refer to forecasts of y_t obtained in this way as *unconditional forecasts* because they are based on forecasts of the exogenous variables for the forecast period. Occasionally, the forecaster may know some or all of the future values of the exogenous variables, for instance, because they are under the control of some decision maker. In that case he or she may be interested in *forecasts* of y_t *conditional* on a specific future path of x_t. In order to derive the optimal conditional forecasts, we write the reduced form (10.2.3) in VARX$(1, 0)$ form,

$$Y_t = \mathbf{A} Y_{t-1} + \mathbf{B} x_t + U_t, \quad (10.5.4)$$

where

$$Y_t := \begin{bmatrix} y_t \\ \vdots \\ y_{t-p+1} \\ x_t \\ \vdots \\ x_{t-s+1} \end{bmatrix}, \quad U_t := \begin{bmatrix} u_t \\ 0 \\ \vdots \\ 0 \end{bmatrix} \; ((Kp+Ms)\times 1),$$

$$\mathbf{A} := \underset{((Kp+Ms)\times(Kp+Ms))}{\left[\begin{array}{cccc|cccc} A_1 & \dots & A_{p-1} & A_p & B_1 & \dots & B_{s-1} & B_s \\ I_K & & 0 & 0 & 0 & \dots & 0 & 0 \\ & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & \dots & I_K & 0 & 0 & \dots & 0 & 0 \\ \hline & & & & 0 & \dots & 0 & 0 \\ & & 0 & & I_M & & 0 & 0 \\ & & & & & \ddots & \vdots & \vdots \\ & & & & 0 & \dots & I_M & 0 \end{array}\right]},$$

and

$$\mathbf{B} := \begin{bmatrix} B_0 \\ 0 \\ \vdots \\ 0 \\ I_M \\ 0 \\ \vdots \\ 0 \end{bmatrix} \begin{matrix} \left.\begin{matrix} \\ \\ \\ \\ \end{matrix}\right\} (Kp \times M) \\ \left.\begin{matrix} \\ \\ \\ \\ \end{matrix}\right\} (Ms \times M) \end{matrix}$$

Successive substitution for lagged Y_t's gives

$$Y_t = \mathbf{A}^h Y_{t-h} + \sum_{i=0}^{h-1} \mathbf{A}^i \mathbf{B} x_{t-i} + \sum_{i=0}^{h-1} \mathbf{A}^i U_{t-i}. \tag{10.5.5}$$

Hence, premultiplying by the $(K \times (Kp+Ms))$ matrix $J := [I_K : 0 : \cdots : 0]$ results in

$$y_{t+h} \;=\; J\mathbf{A}^h Y_t + \sum_{i=0}^{h-1} J\mathbf{A}^i \mathbf{B} x_{t+h-i} + \sum_{i=0}^{h-1} J\mathbf{A}^i J' u_{t+h-i}, \tag{10.5.6}$$

where $U_t = J'JU_t = J'u_t$, has been used. Now the optimal h-step forecast of y_t at origin t, given $x_{t+1}, \dots, x_{t+h}$, and all present and past information, is easily seen to be

$$y_t(h|x) := J\mathbf{A}^h Y_t + \sum_{i=0}^{h-1} J\mathbf{A}^i \mathbf{B} x_{t+h-i} \tag{10.5.7}$$

and the corresponding forecast error is

$$y_{t+h} - y_t(h|x) = \sum_{i=0}^{h-1} J\mathbf{A}^i J' u_{t+h-i}. \tag{10.5.8}$$

Thus, the MSE of the conditional forecast is

$$\Sigma_y(h|x) := \text{MSE}[y_t(h|x)] = \sum_{i=0}^{h-1} J\mathbf{A}^i J' \Sigma_u J(\mathbf{A}^i)' J'. \tag{10.5.9}$$

Although this MSE matrix formally looks like the MSE matrix of the optimal forecast from a VAR model, where $J\mathbf{A}^i J'$ is replaced by Φ_i, the MSE matrix in (10.5.9) is in general different from the one of an unconditional forecast. This fact is easy to see by considering the different definition of the matrix $\mathbf{A}$ used in the pure VAR(p) case.

To illustrate the difference between conditional and unconditional forecasts, we consider the simple reduced form

$$y_t = A_1 y_{t-1} + B_0 x_t + u_t, \tag{10.5.10}$$

where x_t is assumed to be generated by a zero mean VAR(1) process,

$$x_t = C_1 x_{t-1} + v_t.$$

Moreover, we assume that u_t and v_t are independent white noise processes with covariance matrices Σ_u and Σ_v, respectively. The unconditional forecasts are obtained from the VAR process

$$\begin{bmatrix} I_K & -B_0 \\ 0 & I_M \end{bmatrix} \begin{bmatrix} y_t \\ x_t \end{bmatrix} = \begin{bmatrix} A_1 & 0 \\ 0 & C_1 \end{bmatrix} \begin{bmatrix} y_{t-1} \\ x_{t-1} \end{bmatrix} + \begin{bmatrix} u_t \\ v_t \end{bmatrix}$$

which, upon premultiplying with

$$\begin{bmatrix} I_K & -B_0 \\ 0 & I_M \end{bmatrix}^{-1} = \begin{bmatrix} I_K & B_0 \\ 0 & I_M \end{bmatrix},$$

has the standard VAR(1) from

$$\begin{bmatrix} y_t \\ x_t \end{bmatrix} = \begin{bmatrix} A_1 & B_0 C_1 \\ 0 & C_1 \end{bmatrix} \begin{bmatrix} y_{t-1} \\ x_{t-1} \end{bmatrix} + \begin{bmatrix} u_t + B_0 v_t \\ v_t \end{bmatrix}.$$

The optimal 1-step forecast from this model is

$$\begin{bmatrix} y_t(1) \\ x_t(1) \end{bmatrix} = \begin{bmatrix} A_1 & B_0 C_1 \\ 0 & C_1 \end{bmatrix} \begin{bmatrix} y_t \\ x_t \end{bmatrix}.$$

The corresponding MSE matrix is

$$\begin{aligned}\Sigma_*(1) &= E\left(\begin{bmatrix} u_t + B_0 v_t \\ v_t \end{bmatrix}[(u_t + B_0 v_t)', v_t']\right) \\ &= \begin{bmatrix} \Sigma_u + B_0 \Sigma_v B_0' & B_0 \Sigma_v \\ \Sigma_v B_0' & \Sigma_v \end{bmatrix}.\end{aligned}$$

The upper left-hand corner block of this matrix is the MSE matrix of $y_t(1)$, the unconditional forecast of the endogenous variables. Thus,

$$\Sigma_y(1) = \Sigma_u + B_0 \Sigma_v B_0'. \tag{10.5.11}$$

On the other hand, in the VARX(1,0) representation (10.5.4), we have $\mathbf{A} = A_1$ and $\mathbf{B} = B_0$ for the present example. Hence, the conditional 1-step forecast of y_t is

$$y_t(1|x) = A_1 y_t + B_0 x_{t+1}$$

with corresponding MSE matrix

$$\Sigma_y(1|x) = \Sigma_u.$$

Obviously, $\Sigma_y(1) - \Sigma_y(1|x) = B_0 \Sigma_v B_0'$ is positive semidefinite and, thus, the unconditional forecast is inferior to the conditional forecast, if $B_0 \neq 0$. It must be kept in mind, however, that the conditional forecast is only feasible if the future values of the exogenous variables are either known or assumed. If only hypothetical values are used, the conditional forecast may be quite poor if the actual values of the exogenous variables turn out to be different from the hypothetical ones. The smaller MSE of the conditional forecast is simply due to ignoring any uncertainty regarding the future paths of the exogenous variables.

Using the foregoing results, interval forecasts and forecast regions can be set up as usual. It may also be worth pointing out that we have not used the stability of the VAR operator or stationarity of the variables. Hence, the formulas are also valid for systems with integrated and cointegrated variables. So far we have discussed forecasting with known models. The case of estimated models will be considered next.

10.5.2 Forecasting Estimated Dynamic SEMs

In order to evaluate the consequences of using estimated instead of known processes for unconditional forecasts, we can use a joint model for the endogenous and exogenous variables and then draw on results of the previous chapters. Therefore, in this section we will focus on conditional forecasts only. We denote by $\widehat{y}_t(h|x)$ the conditional h-step forecast (10.5.7) based on the estimated reduced form (10.2.3). The forecast error is

$$y_{t+h} - \widehat{y}_t(h|x) = [y_{t+h} - y_t(h|x)] + [y_t(h|x) - \widehat{y}_t(h|x)]. \tag{10.5.12}$$

Conditional on the exogenous variables, the two terms in brackets are uncorrelated. Hence, assuming, as in previous chapters, that the processes used for estimation and forecasting are independent, an MSE approximation

$$\Sigma_{\widehat{y}}(h|x) = \Sigma_y(h|x) + \frac{1}{T}\Omega_y(h|x) \tag{10.5.13}$$

is obtained in the by now familiar way. Here

$$\Omega_y(h|x) := E\left[\frac{\partial y_t(h|x)}{\partial \boldsymbol{\beta}'}\Sigma_{\widehat{\boldsymbol{\beta}}}\frac{\partial y_t(h|x)'}{\partial \boldsymbol{\beta}}\right], \tag{10.5.14}$$

$\boldsymbol{\beta} := \mathrm{vec}[A_1, \dots, A_p, B_1, \dots, B_s, B_0]$ and $\Sigma_{\widehat{\boldsymbol{\beta}}}$ is the covariance matrix of the asymptotic distribution of $\sqrt{T}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta})$. It is straightforward to show that

$$\begin{aligned}\frac{\partial y_t(h|x)}{\partial \boldsymbol{\beta}'} &= \frac{\partial(J\mathbf{A}^h Y_t)}{\partial \boldsymbol{\beta}'} + \sum_{i=0}^{h-1}\frac{\partial(J\mathbf{A}^i\mathbf{B}x_{t+h-i})}{\partial \boldsymbol{\beta}'} \\ &= \sum_{i=0}^{h-1}\Big[Y_t'(\mathbf{A}')^{h-1-i} \otimes J\mathbf{A}^i J' \\ &\qquad + \sum_{j=0}^{i-1} x_{t+h-i}'\mathbf{B}'(\mathbf{A}')^{i-1-j} \otimes J\mathbf{A}^j J' : x_{t+h-i}' \otimes J\mathbf{A}^i J'\Big].\end{aligned} \tag{10.5.15}$$

For stationary processes, an estimator of $\Omega_y(h|x)$ is obtained in the usual way be replacing all unknown parameters in this expression and in $\Sigma_{\widehat{\boldsymbol{\beta}}}$ by estimators and by using the average over $t = 1, \dots, T$ for the expectation in (10.5.14).

Although we have discussed forecasting with estimated coefficients in terms of a simple VARX(p, s) model with white noise residuals, it is possible to generalize these results to models with autocorrelated error processes. The more general case was treated, for instance, by Yamamoto (1980) and Baillie (1981).

10.6 Multiplier Analysis

In an econometric simultaneous equations analysis, the marginal impact of changes in the exogenous variables is sometimes investigated. For example, if the exogenous variables are instruments for, say, the government or a central bank the consequences of changes in these instruments may be of interest. A government may, for instance, desire to know the effects of a change in a tax rate. In that case, *policy simulation* is of interest. In other cases, the consequences of changes in the exogenous variables that are not under the control of any decision maker may be of interest. For instance, it may be desirable to study the future consequences of the present weather conditions.

Therefore, the dynamic multipliers discussed in Section 10.2.2 are considered. They are contained in the D_i matrices of the final form operator,

$$D(L) = \sum_{i=0}^{\infty} D_i L^i := A(L)^{-1} B(L),$$

where $A(L) := I_K - A_1 L - \cdots - A_p L^p$ and $B(L) := B_0 + B_1 L + \cdots + B_s L^s$ are the reduced form operators, as before. Here stability and, hence, invertibility of the VAR operator $A(L)$ is assumed. The D_i matrices are conveniently obtained from the VARX$(1,0)$ representation (10.5.4) which implies

$$y_t = \sum_{i=0}^{\infty} J\mathbf{A}^i \mathbf{B} x_{t-i} + \sum_{i=0}^{\infty} J\mathbf{A}^i J' u_{t-i}, \tag{10.6.1}$$

because $J\mathbf{A}^h Y_t \to 0$ as $h \to \infty$, if y_t is a stable, stationary process (see (10.5.6)). The D_i's are coefficient matrices of the exogenous variables in the final form representation. Thus,

$$D_i = J\mathbf{A}^i \mathbf{B}, \quad i = 0, 1, \ldots, \tag{10.6.2}$$

the n-th interim multipliers are

$$M_n := D_0 + D_1 + \cdots + D_n = J(I + \mathbf{A} + \cdots + \mathbf{A}^n)\mathbf{B}, \quad n = 0, 1, \ldots, \tag{10.6.3}$$

and the total multipliers are

$$M_\infty := \sum_{i=0}^{\infty} D_i = J(I - \mathbf{A})^{-1}\mathbf{B} = A(1)^{-1} B(1). \tag{10.6.4}$$

If the model contains integrated variables and the generation mechanism is started at time $t = 0$, say, from a set of initial values, then we get from (10.5.5),

$$y_t = J\mathbf{A}^t Y_0 + \sum_{i=0}^{t-1} J\mathbf{A}^i \mathbf{B} x_{t-i} + \sum_{i=0}^{t-1} J\mathbf{A}^i J' u_{t-i}. \tag{10.6.5}$$

Thus, the D_i matrices in (10.6.2) still reflect the marginal impacts of changes in the unmodelled variables and, hence, contain the multipliers. Also the n-th interim multipliers can be computed as in (10.6.3), whereas the total multipliers in (10.6.4) will not exist in general.

Having obtained the foregoing representations of the multipliers, estimation of these quantities is straightforward. Estimators of the dynamic multipliers are obtained by substituting estimators $\widehat{A}_i$ and $\widehat{B}_j$ of the coefficient matrices in $\mathbf{A}$ and $\mathbf{B}$. The asymptotic properties of the estimators then follow in the usual way. For completeness we mention the following result from Schmidt (1973).

In the framework of Section 10.3, suppose $\widehat{\boldsymbol{\beta}}$ is a consistent estimator of $\boldsymbol{\beta} :=$ vec$[A, B, B_0]$ satisfying

$$\sqrt{T}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \xrightarrow{d} \mathcal{N}(0, \Sigma_{\widehat{\boldsymbol{\beta}}}).$$

Then

$$\sqrt{T}\,\text{vec}(\widehat{D}_i - D_i) \xrightarrow{d} \mathcal{N}(0, G_i \Sigma_{\widehat{\boldsymbol{\beta}}} G_i'), \tag{10.6.6}$$

where $G_0 := [0 : I_{KM}]$ and

$$G_i := \frac{\partial\,\text{vec}(D_i)}{\partial \boldsymbol{\beta}'} = \left[\sum_{j=0}^{i-1} \mathbf{B}'(\mathbf{A}')^{i-1-j} \otimes J\mathbf{A}^j J' : I_M \otimes J\mathbf{A}^i J' \right], \quad i = 1, 2, \ldots,$$

are $[KM \times (K^2 p + KM(s+1))]$ matrices. The proof of this result is left as an exercise. It is also easy to find the asymptotic distribution of the interim multipliers (accumulated multipliers) and the total multipliers if they exist (see Problem 10.8).

10.7 Optimal Control

A policy or decision maker who has control over some of the exogenous variables can use a dynamic simultaneous equations model to assess interventions with a multiplier or simulation analysis, as described in the previous section. However, if the decision maker has specific target values of the endogenous variables in mind, he or she may wish to go a step further and determine which values of the instrument variables will produce the desired values of the endogenous variables.

Usually it will not be possible to actually achieve all targets simultaneously and sometimes the decision maker is not completely free to choose the instruments. For instance, doubling a particular tax rate or increasing the price of specific government services drastically may result in the overthrow of the government or in social unrest and is therefore not a feasible option. Therefore, a loss function is usually set up in which the loss of deviations from the target values is specified. For instance, if the desired paths of the endogenous and instrument variables after period T are $y^0_{T+1}, \ldots, y^0_{T+n}$ and $x^0_{T+1}, \ldots, x^0_{T+n}$, respectively, a *quadratic* loss function has the form

$$\begin{aligned} \mathfrak{L} \;=\; & \sum_{i=1}^{n} [(y_{T+i} - y^0_{T+i})' K_i (y_{T+i} - y^0_{T+i}) \\ & \qquad + (x_{T+i} - x^0_{T+i})' P_i (x_{T+i} - x^0_{T+i})], \end{aligned} \tag{10.7.1}$$

where the K_i and P_i are symmetric positive semidefinite matrices. Because the variables are assumed to be stochastic, the loss is a random variable too.

Therefore, minimization of the average or expected loss, $E(\mathfrak{L})$, is usually the objective.

In a quadratic loss function the same weight is assigned to positive and negative deviations from the target values. For many situations and variables this specification is not quite realistic. For example, if the target is to have an unemployment rate of 2%, then having less than 2% may not be a problem at all while any higher rate may be regarded as a serious problem. Nevertheless, quadratic loss functions are the most common ones in applied and theoretical studies. Therefore, we will also use them in the following. One reason for the popularity of this type of loss function is clearly its tractability.

In order to approach a formal solution of the optimal control problem outlined in the foregoing, we assume that the economic system is described by a model like (10.1.1) with reduced form (10.2.3). However, to be able to distinguish between instrument variables and other exogenous variables, we introduce a new symbol for the latter. Suppose x_t represents an $(M \times 1)$ vector of instrument variables, the $(N \times 1)$ vector z_t contains all other unmodelled variables and the reduced form of the model is

$$y_t = A_1 y_{t-1} + \cdots + A_p y_{t-p} + B_0 x_t + \cdots + B_s x_{t-s} + C z_t + u_t, \qquad (10.7.2)$$

where u_t is white noise. Some of the components of z_t may be lagged variables. To summarize them in a vector indexed by t is just a matter of convenience.

For the present purposes, it is useful to write the model in VARX$(1, 0)$ form similar to (10.5.4),

$$Y_t = \mathbf{A} Y_{t-1} + \mathbf{B} x_t + \mathbf{C} z_t + U_t, \qquad (10.7.3)$$

where $Y_t, U_t, \mathbf{A}$, and $\mathbf{B}$ are as defined in (10.5.4) and

$$\mathbf{C} := \begin{bmatrix} C \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

is a $((Kp + Ms) \times N)$ matrix. Recall that

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

contains current and lagged endogenous and instrument variables. Thus, the quadratic loss function specified in (10.7.1) may be rewritten in the form

$$\mathfrak{L} = \sum_{i=1}^{n} (Y_{T+i} - Y_{T+i}^0)' Q_i (Y_{T+i} - Y_{T+i}^0), \tag{10.7.4}$$

where the Q_i are symmetric positive semidefinite matrices involving the K_i's and P_i's.

In this framework, the problem of *optimal control* may be stated as follows: Given the model (10.7.3), given the vector Y_T, given values $z_{T+1}, \dots, z_{T+n}$ of the uncontrolled variables and given target values $y_{T+1}^0, \dots, y_{T+n}^0$ and $x_{T+1}^0, \dots, x_{T+n}^0$, find the values $x_{T+1}^*, \dots, x_{T+n}^*$ that minimize the expected loss $E(\mathfrak{L})$ specified in (10.7.4). The solution to this dynamic programming problem is well documented in the control theory literature. It turns out to be

$$x_{T+i}^* = G_i Y_{T+i-1} + g_i, \quad i = 1, \dots, n, \tag{10.7.5}$$

where the Y_{T+i} are assumed to be obtained as

$$Y_{T+i} = \mathbf{A} Y_{T+i-1} + \mathbf{B} x_{T+i}^* + \mathbf{C} z_{T+i} + u_{T+i}.$$

Here the $(M \times (Kp + Ms))$ matrix G_i is defined as

$$G_i := -(\mathbf{B}' H_i \mathbf{B})^{-1} \mathbf{B}' H_i \mathbf{A}$$

and the $(M \times 1)$ vector g_i is defined as

$$g_i := -(\mathbf{B}' H_i \mathbf{B})^{-1} \mathbf{B}' (H_i \mathbf{C} z_{T+i} - h_i)$$

with

$$H_n := Q_n \quad \text{and} \quad H_{i-1} := Q_{i-1} + (\mathbf{A} + \mathbf{B} G_i)' H_i (\mathbf{A} + \mathbf{B} G_i), \quad \text{for } i = 1, \dots, n-1,$$

and

$$h_n := Q_n Y_{T+n}^0 \quad \text{and}$$
$$h_{i-1} := Q_{i-1} Y_{T+i-1}^0 - \mathbf{A}' H_i (\mathbf{C} z_{T+i} + \mathbf{B} g_i) + \mathbf{A}' h_i \quad \text{for } i = 1, \dots, n-1.$$

The actual computation of these quantities proceeds in the order H_n, G_n, h_n, g_n, H_{n-1}, G_{n-1}, h_{n-1}, g_{n-1}, H_{n-2}, ... This solution can be found in various variations in the control theory literature (e.g., Chow (1975, 1981), Murata (1982)). Obviously, because the Y_t are random, the same is true for the optimal decision rule x_{T+i}^*, $i = 1, \dots, n$.

There are a number of problems that arise in practice in the context of optimal control as presented here. For instance, we have considered a finite planning horizon of n periods. In some situations it is of interest to find the optimal decision rule for an infinite planning period. Moreover, in practice the parameter matrices $\mathbf{A}$, $\mathbf{B}$, and $\mathbf{C}$ are usually unknown and have to be replaced by estimators. More generally, stochastic parameter models may be considered. This, of course, introduces an additional stochastic element into the optimal decision rule. A further complication arises if the relations between the variables cannot be captured adequately by a *linear* model such as (10.7.2) but require a nonlinear specification. It is also possible to consider other types of optimization rules. In this section, we have assumed that the optimal decision rule for period $T + i$ is determined on the basis of all available information in period $T + i - 1$. In particular, the realization Y_{T+i-1} is assumed to be given in setting up the decision rule x^*_{T+i}. Such an approach is often referred to as a *closed-loop strategy*. An alternative approach would be to determine the decision rule at the beginning of the planning period for the entire planning horizon. This approach is called an *open-loop strategy*. Although it is in general inferior to closed-loop optimization, it may be of interest occasionally. These and many other topics are treated in the optimal control literature. Chow (1975, 1981) and Murata (1982) are books on the topic with emphasis on optimal decision making related to economic and econometric models. Friedmann (1981) provided the asymptotic properties of the optimal decision rule when estimators are substituted for the parameters in the control rule.

10.8 Concluding Remarks on Dynamic SEMs

In this chapter, we have summarized some problems related to the estimation, specification, and analysis of dynamic models with unmodelled variables. Major problem areas that were identified without giving details of possible solutions are the distinction between endogenous and exogenous variables, the identification or unique parameterization of dynamic models, the estimation, specification, and checking of structural form models as well as the treatment of nonlinear specifications. Also, we have just scratched the surface of control problems which represent one important area of applications of dynamic SEMs.

Other problems of obvious importance in the context of these models relate to the choice of the data associated with the variables. If a structural form is derived from some economic or other subject matter theory, it is important that the available data represents realizations of the variables related to the theory. In particular, the level of aggregation (temporal and contemporaneous) and seasonal characteristics (seasonally adjusted or unadjusted) may be of importance. The models we have considered do not allow specifically for seasonality, except perhaps for seasonal dummies and other seasonal components among the unmodelled variables. The seasonality aspect in the context of dynamic SEMs and models specifically designed for seasonal data were discussed, for example, by Hylleberg (1986).

So far, we have essentially considered stationary and integrated processes. Mild deviations from the stationarity assumption are possible in dynamic

SEMs where unmodelled variables may cause changes in the mean or conditional mean of the endogenous variables. However, in discussing properties of estimators or long-run multipliers, we have made assumptions that come close to assuming stationarity or cointegration. For instance, if the unmodelled variables are driven by a stationary VAR process, the means and second moments of the endogenous variables may be time invariant. Unfortunately, in practice, changes in the data generation process may occur. Therefore, we will discuss specific types of models with time varying parameters in later chapters (see Chapters 17 and 18).

10.9 Exercises

Problem 10.1
Consider the following structural form

$$\begin{aligned} Q_t &= \alpha_0 + \alpha_1 R_{t-1} + w_{1t}, \\ P_t &= \beta_0 + \beta_1 Q_t + w_{2t}, \end{aligned}$$

where R_t is a measure for the rainfall in period t, Q_t is the quantity of an agricultural product supplied in period t, and P_t is the price of the product. Derive the reduced form, the final equations, and the final form of the model.

Problem 10.2
Suppose that the rainfall variable R_t in Problem 10.1 is generated by a white noise process with mean μ_R. Determine the unconditional 3-step ahead forecasts for Q_t and P_t based on the model from Problem 10.1. Determine also the conditional 3-step ahead forecasts given $R_{t+i} = \mu_R$, $i = 1, 2, 3$. Compare the two forecasts.

Problem 10.3
Given the model of Problem 10.1, what is the marginal total or long-run effect of an additional unit of rainfall in period t?

Problem 10.4
Suppose the system y_t has the structural form

$$A^*(L)y_t = F^* y_t^e + B^*(L)x_t + w_t,$$

where $A^*(L) := \mathsf{A} - A_1^* L - \cdots - A_p^* L^p$, $B^*(L) := B_0^* + B_1^* L + \cdots + B_s^* L^s$ and x_t is generated by a VAR(q) process

$$C(L)x_t = v_t.$$

Assume that y_t^e represents rational expectations formed in period $t-1$ and eliminate the expectations variables from the structural form.

Problem 10.5
Show that the 1-step ahead forecast for y_t obtained from the VAR(p) model (10.5.3) is identical to the one determined from (10.5.2) if

$$x_t(1) = C_1 x_t + \cdots + C_q x_{t-q+1}$$

is used as forecast for the exogenous variables.

Problem 10.6
Show that the partial derivatives $\partial y_t(h|x)/\partial \boldsymbol{\beta}'$ have the form given in (10.5.15).

Problem 10.7
Derive a prediction test for structural change on the basis of the conditional forecasts of the endogenous variables of a dynamic SEM.

Problem 10.8
Show that the dynamic multipliers have the asymptotic distributions given in Section 10.6. Show also that the n-th interim multipliers have an asymptotic normal distribution,

$$\sqrt{T}\,\mathrm{vec}(\widehat{M}_n - M_n) \xrightarrow{d} \mathcal{N}(0, \Sigma_{\widehat{\mathbf{m}}}(n)),$$

where

$$\Sigma_{\widehat{\mathbf{m}}}(n) = (G_0 + \cdots + G_n)\Sigma_{\widehat{\boldsymbol{\beta}}}(G_0 + \cdots + G_n)'$$

and the G_i are the $[KM \times K(Kp + M(s+1))]$ matrices defined in Section 10.6. Furthermore,

$$\sqrt{T}\,\mathrm{vec}(\widehat{M}_\infty - M_\infty) \xrightarrow{d} \mathcal{N}(0, \Sigma_{\widehat{\mathbf{m}}}(\infty)),$$

where

$$\Sigma_{\widehat{\mathbf{m}}}(\infty) = G_\infty \Sigma_{\widehat{\boldsymbol{\beta}}} G_\infty'$$

with

$$G_\infty := [((I - \mathbf{A})^{-1}\mathbf{B})' : I_M] \otimes J(I - \mathbf{A})^{-1}J'.$$

Here the notation from Section 10.6 is used.

Problem 10.9
Derive the optimal decision rule for the control problem stated in Section 10.7. (Hint: See Chow (1975).)