6

Vector Error Correction Models

As defined in Chapter 2, a process is stationary if it has time invariant first and second moments. In particular, it does not have trends or changing variances. A VAR process has this property if the determinantal polynomial of its VAR operator has all its roots outside the complex unit circle. Clearly, stationary processes cannot capture some main features of many economic time series. For example, trends (trending means) are quite common in practice. For instance, the *original* investment, income, and consumption data used in many previous examples have trends (see Figure 3.1). Thus, if interest centers on analyzing the original variables (or their logarithms) rather than the rates of change, it is necessary to have models that accommodate the nonstationary features of the data. It turns out that a VAR process can generate stochastic and deterministic trends if the determinantal polynomial of the VAR operator has roots on the unit circle. In fact, it is even sufficient to allow for unit roots (roots for $z = 1$) to obtain a trending behavior of the variables. We will consider this case in some detail in this chapter. In the next section, the effect of unit roots in the AR operator of a univariate process will be analyzed. Variables generated by such processes are called *integrated variables* and the underlying generating processes are *integrated processes*. Vector processes with unit roots are considered in Section 6.2. In these processes, some of the variables can have common trends so that they move together to some extent. They are then called *cointegrated*. This feature is considered in detail in Section 6.3 and it is shown that *vector error correction models* (*VECMs*) offer a convenient way to parameterize and specify them. In Section 6.3, the processes are assumed to be purely stochastic and do not have deterministic terms. How to incorporate these terms is the subject of Section 6.4. Once we have a suitable model setup, it can be used for forecasting, causality analysis, and impulse response analysis. These issues are treated in Sections 6.5–6.7.

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

6.1 Integrated Processes

Recall that a VAR(p) process,

$$y_t = A_1 y_{t-1} + \cdots + A_p y_{t-p} + u_t, \tag{6.1.1}$$

is stable if the polynomial defined by

$$\det(I_K - A_1 z - \cdots - A_p z^p)$$

has no roots in and on the complex unit circle. For a univariate AR(1) process, $y_t = \alpha y_{t-1} + u_t$, this property means that

$$1 - \alpha z \neq 0 \quad \text{for } |z| \leq 1$$

or, equivalently, $|\alpha| < 1$.

Consider the borderline case, where $\alpha = 1$. The resulting process $y_t = y_{t-1} + u_t$ is called a *random walk*. Starting the process at $t = 0$ with some fixed y_0, it is easy to see by successive substitution for lagged y_t's, that

$$y_t = y_{t-1} + u_t = y_{t-2} + u_{t-1} + u_t = \cdots = y_0 + \sum_{i=1}^{t} u_i. \tag{6.1.2}$$

Thus, y_t consists of the sum of all disturbances or innovations of the previous periods so that each disturbance has a lasting impact on the process. If u_t is white noise with variance σ_u^2,

$$E(y_t) = y_0$$

and

$$\text{Var}(y_t) = t\text{Var}(u_t) = t\sigma_u^2.$$

Hence, the variance of a random walk tends to infinity. Furthermore, the correlation

$$\begin{aligned}\text{Corr}(y_t, y_{t+h}) &= \frac{E\left[\left(\sum_{i=1}^{t} u_i\right)\left(\sum_{i=1}^{t+h} u_i\right)\right]}{[t\sigma_u^2(t+h)\sigma_u^2]^{1/2}} \\ &= \frac{t}{(t^2+th)^{1/2}} \underset{t\to\infty}{\longrightarrow} 1\end{aligned}$$

for any integer h. This latter property of a random walk means that y_t and y_s are strongly correlated even if they are far apart in time. It can also be shown that the expected time between two crossings of zero is infinite. These properties are often reflected in trending behavior. Examples are depicted in Figure 6.1. This kind of trend is, of course, not a deterministic one but a stochastic trend.

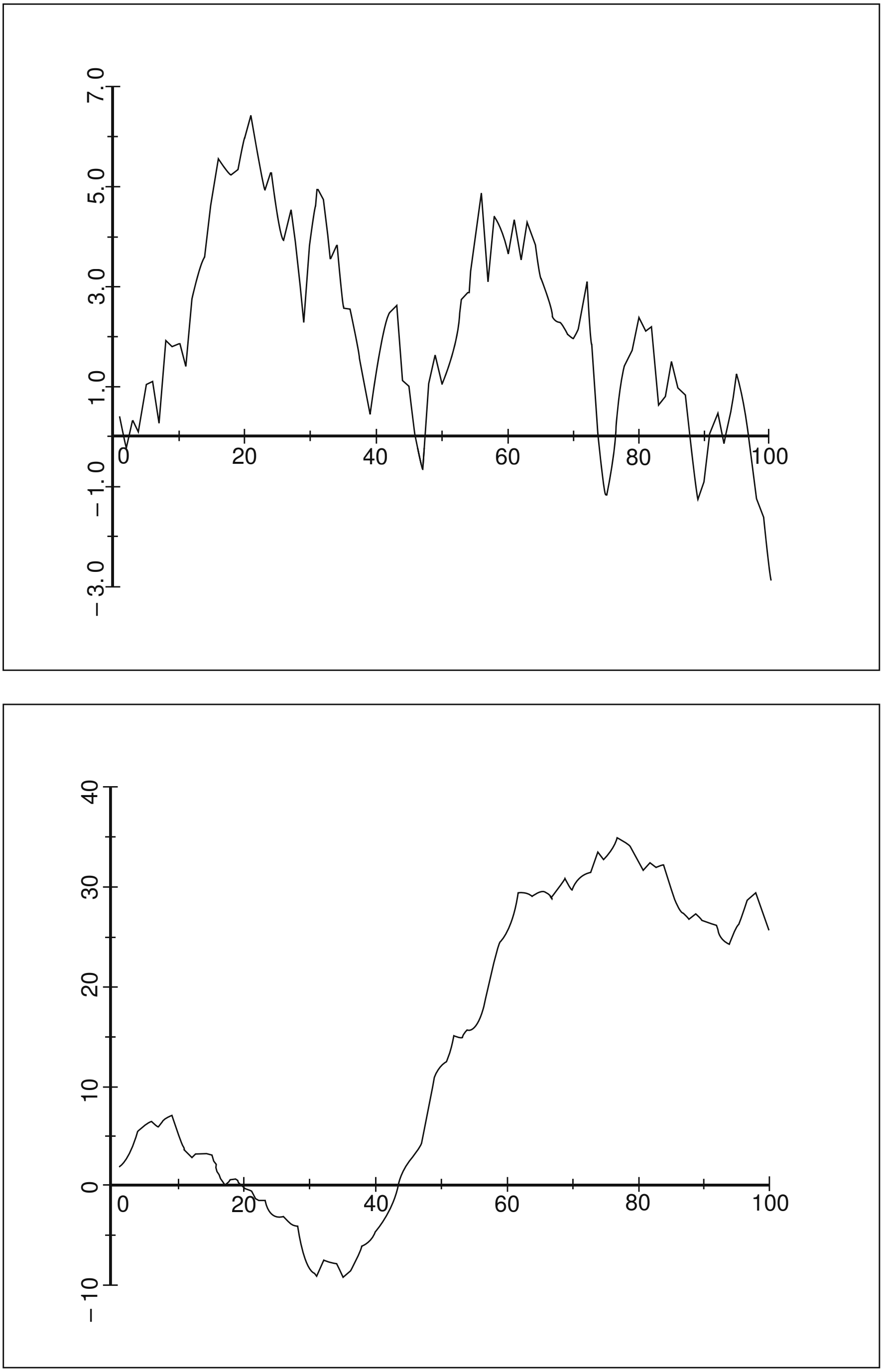

Fig. 6.1. Artificially generated random walks.

If the process has a nonzero constant term ν, $y_t = \nu + y_{t-1} + u_t$ is called a *random walk with drift* and it has a deterministic linear trend in the mean. To see this property, suppose again that the process is started at $t = 0$ with a fixed y_0. Then

$$y_t = y_0 + t\nu + \sum_{i=1}^{t} u_i$$

and $E(y_t) = y_0 + t\nu$. An example of a time series generated by a random walk with drift is shown in Figure 6.2.

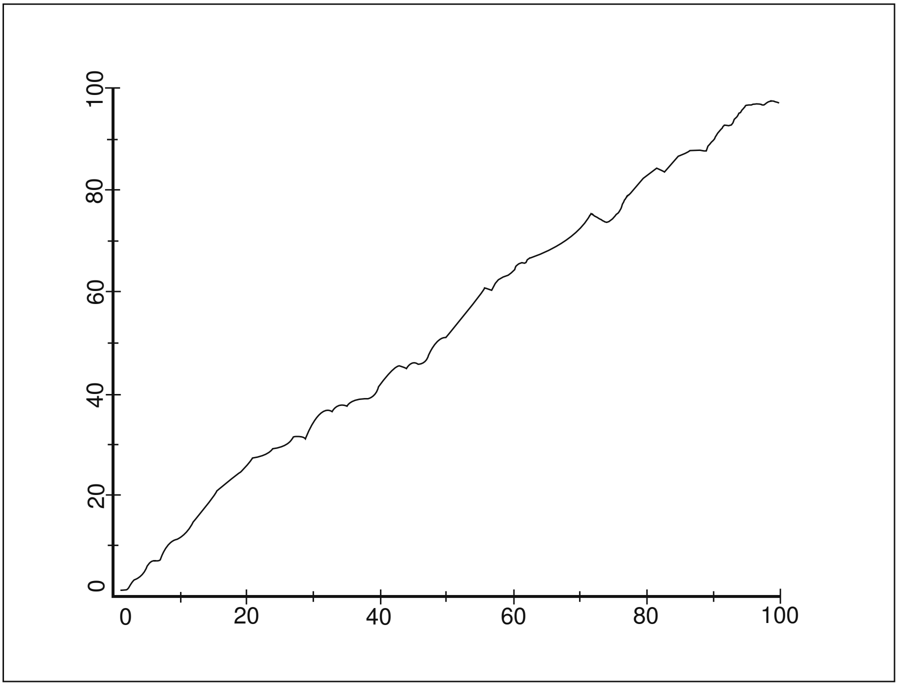

Fig. 6.2. An artificially generated random walk with drift.

The previous discussion suggests that starting unstable processes at some finite time t_0 is useful to obtain processes with finite moments. On the other hand, if an AR process starts at some finite time, it is strictly speaking not necessarily stationary, even if it is stable. To see this property, let $y_t = \nu + \alpha y_{t-1} + u_t$ be a univariate stable AR(1) process with $|\alpha| < 1$. Starting with a random variable y_0 at $t = 0$, gives

$$y_t = \nu \sum_{i=0}^{t-1} \alpha^i + \alpha^t y_0 + \sum_{i=0}^{t-1} \alpha^i u_{t-i}.$$

Hence,

$$E(y_t) = \nu \sum_{i=0}^{t-1} \alpha^i + \alpha^t E(y_0)$$

is generally not time invariant if α and $\nu \neq 0$. A similar result is obtained for the second moments,

$$\mathrm{Var}(y_t) = \alpha^{2t}\mathrm{Var}(y_0) + \sigma_u^2 \sum_{i=0}^{t-1} \alpha^{2i}.$$

However, the first and second moments approach limit values as $t \to \infty$ and one might call such a process *asymptotically stationary*. To simplify matters, the term "asymptotically" is sometimes dropped and such processes are then simply called stationary. Moreover, if we consider purely stochastic processes without deterministic terms ($\nu = 0$), the initial variable can be chosen such that y_t is stationary if the process is stable. In particular, if we choose

$$y_0 = \sum_{i=0}^{\infty} \alpha^i u_{-i}$$

we get, for $\nu = 0$,

$$y_t = \alpha^t \sum_{i=0}^{\infty} \alpha^i u_{-i} + \sum_{i=0}^{t-1} \alpha^i u_{t-i} = \sum_{i=0}^{\infty} \alpha^i u_{t-i}, \quad t = 1, 2, \dots,$$

and, hence, for $t = 1, 2, \dots,$

$$E(y_t) = 0,$$

$$\mathrm{Var}(y_t) = \sigma_u^2/(1 - \alpha^2),$$

and also the autocovariances are time invariant. Thus, for a stable process we may in fact choose the initial variable such that y_t is stationary even if the process is started in some given period. This result can also be used as a justification for simply calling stable processes stationary in this situation. We may implicitly assume that the starting value is chosen to justify the terminology. For our purposes, this point is of limited importance because in later chapters we will be interested in the parameters of the processes considered and possibly in their asymptotic moments. Without further warning, nonstationary, unstable processes will be assumed to begin at some given finite time period.

A behavior similar to that of a random walk is also observed for higher order AR processes such as

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

if $1 - \alpha_1 z - \cdots - \alpha_p z^p$ has a root for $z = 1$. Note that

$$1 - \alpha_1 z - \cdots - \alpha_p z^p = (1 - \lambda_1 z) \cdots (1 - \lambda_p z),$$

where $\lambda_1, \dots, \lambda_p$ are the reciprocals of the roots of the polynomial. If the process has just one *unit root* (a root equal to 1) and all other roots are outside the complex unit circle, its behavior is similar to that of a random walk, that is, its variances increase linearly, the correlation between variables h periods apart tends to 1 and the process has a linear trend in mean if $\nu \neq 0$. In case one of the roots is strictly inside the unit circle, the process becomes explosive, that is, its variances go to infinity at an exponential rate. Many researchers feel that such processes are unrealistic models for most economic data. Although processes with roots on the unit circle other than one are often useful, we shall concentrate on the case of unit roots and all other roots outside the unit circle. This situation is of considerable practical interest.

Univariate processes with d unit roots (d roots equal to 1) in their AR operators are called *integrated of order d* ($I(d)$). If there is just one unit root, i.e., the process is $I(1)$, it is quite easy to see how a stable and possibly stationary process can be obtained: simply by taking first differences, $\Delta y_t := (1 - L) y_t = y_t - y_{t-1}$, of the original process. More generally, if the process is $I(d)$ it can be made stable by differencing d times, that is, $\Delta^d y_t = (1 - L)^d y_t$ is stable and, again, initial values can be chosen such that it is stationary. In the following, it will often be convenient to extend this terminology also to stable, stationary processes and to call them $I(0)$.

More generally, y_t may be defined to be an $I(1)$ process, if $\Delta y_t = w_t$ is a stationary process with infinite MA representation, $w_t = \sum_{j=0}^{\infty} \theta_j u_{t-j} = \theta(L) u_t$, where the MA coefficients satisfy the condition $\sum_{j=0}^{\infty} j|\theta_j| < \infty$, $\theta(1) = \sum_{j=0}^{\infty} \theta_j \neq 0$, and $u_t \sim (0, \sigma_u^2)$ is white noise. In that case, $y_t = y_{t-1} + w_t$ can be rewritten as

$$y_t = y_0 + w_1 + \cdots + w_t = y_0 + \theta(1)(u_1 + \cdots + u_t) + \sum_{j=0}^{\infty} \theta_j^* u_{t-j} - w_0^*, \quad (6.1.3)$$

where $\theta_j^* = -\sum_{i=j+1}^{\infty} \theta_i$, $j = 0, 1, \dots$, and $w_0^* = \sum_{j=0}^{\infty} \theta_j^* u_{-j}$ contains initial values. Thus, y_t can be represented as the sum of a random walk $[\theta(1)(u_1 + \cdots + u_t)]$, a stationary process $[\sum_{j=0}^{\infty} \theta_j^* u_{t-j}]$, and initial values $[y_0 - w_0^*]$. Notice that the condition $\sum_{j=0}^{\infty} j|\theta_j| < \infty$ ensures that $\sum_{j=0}^{\infty} |\theta_j^*| < \infty$, so that $\sum_{j=0}^{\infty} \theta_j^* u_{t-j}$ is indeed well-defined by Proposition C.7 of Appendix C.3. Although the condition for the θ_j is stronger than absolute summability, it is satisfied for many processes of practical interest. The decomposition of y_t in (6.1.3) is known as the *Beveridge-Nelson decomposition* (see also Appendix C.8). A similar decomposition for multivariate processes is helpful in some of the subsequent analysis. It will be discussed in Section 6.3.

6.2 VAR Processes with Integrated Variables

Consider now a K-dimensional VAR(p) process without a deterministic term as in (6.1.1). It can be written as

$$A(L)y_t = u_t, \tag{6.2.1}$$

where $A(L) := I_K - A_1 L - \cdots - A_p L^p$ and L is the lag operator. Multiplying from the left by the adjoint $A(L)^{adj}$ of $A(L)$ gives

$$|A(L)|y_t = A(L)^{adj} u_t \tag{6.2.2}$$

(see Appendix A.4.1 for the definition of the adjoint of a matrix). Thus, the VAR(p) process in (6.2.1) can be written as a process with univariate AR operator, that is, all components have the same AR operator $|A(L)|$. The right-hand side of (6.2.2), $A(L)^{adj} u_t$, is a finite order MA process (see Chapter 11 for further discussion of such processes). If $|A(L)|$ has d unit roots and otherwise all roots are outside the unit circle, the AR operator can be written as

$$|A(L)| = \alpha(L)(1 - L)^d = \alpha(L)\Delta^d,$$

where $\alpha(L)$ is an invertible operator. Consequently, $\Delta^d y_t$ is a stable process. Hence, each component becomes stable upon differencing.

Because we are considering processes which are started at some specific time t_0, we should perhaps think for a moment about the treatment of initial values when multiplying by an operator such as $A(L)^{adj}$ in (6.2.2). One possible assumption is that the new representation is valid for all t for which the y_t's are defined in (6.2.1).

The foregoing discussion shows that if a VAR(p) process is unstable because of unit roots only, it can be made stable by differencing its components. Note, however, that, due to cancellations, it may not be necessary to difference each component as many times as there are unit roots in $|A(L)|$. To illustrate this point, consider the bivariate VAR(1) process

$$\left(\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} - \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} L \right) \begin{bmatrix} y_{1t} \\ y_{2t} \end{bmatrix} = \begin{bmatrix} (1-L)y_{1t} \\ (1-L)y_{2t} \end{bmatrix} = u_t.$$

Obviously, each component is stationary after differencing once, i.e., each component is $I(1)$, although

$$|A(L)| = \left| \begin{bmatrix} 1-L & 0 \\ 0 & 1-L \end{bmatrix} \right| = (1-L)^2$$

has two unit roots. It is also possible that some components are stable and stationary as univariate processes whereas others need differencing. Examples are easy to construct.

If the VAR(p) process has a nonzero intercept term so that

$$A(L)y_t = \nu + u_t$$

and $|A(z)|$ has one or more unit roots, then some of the components of y_t may have deterministic trends in their mean values. Unlike the univariate case, it is also possible, however, that none of the components of y_t has a deterministic trend in mean. This occurs if $A(L)^{adj}\nu = 0$. For instance, if

$$A(L) = \begin{bmatrix} 1-L & \eta L \\ 0 & 1 \end{bmatrix},$$

$|A(z)|$ has a unit root and

$$A(L)^{adj} = \begin{bmatrix} 1 & -\eta L \\ 0 & 1-L \end{bmatrix}.$$

Hence,

$$A(L)^{adj} \begin{bmatrix} \nu_1 \\ \nu_2 \end{bmatrix} = \begin{bmatrix} \nu_1 - \eta\nu_2 \\ \nu_2 - \nu_2 \end{bmatrix}$$

which is zero if $\nu_1 = \eta\nu_2$. Thus, in a VAR analysis an intercept term cannot be excluded a priori if there are unit roots and none of the component series has a deterministic trend.

The following question comes to mind in this context. Suppose each component of a VAR(p) process is $I(d)$, is it possible that differencing each component individually distorts interesting features of the relationship between the original variables? If the latter were not the case, a VAR analysis could be performed as described in previous chapters after differencing the individual components. It turns out, however, that differencing may indeed distort the relationship between the original variables. Systems with cointegrated variables are examples, where fitting VAR models upon differencing may be inadequate. Such systems are introduced next.

6.3 Cointegrated Processes, Common Stochastic Trends, and Vector Error Correction Models

Equilibrium relationships are suspected between many economic variables such as household income and expenditures or prices of the same commodity in different markets. Suppose the variables of interest are collected in the vector $y_t = (y_{1t}, \dots, y_{Kt})'$ and their long-run equilibrium relation is $\beta' y_t = \beta_1 y_{1t} + \cdots + \beta_K y_{Kt} = 0$, where $\beta = (\beta_1, \dots, \beta_K)'$. In any particular period, this relation may not be satisfied exactly but we may have $\beta' y_t = z_t$, where z_t is a stochastic variable representing the deviations from the equilibrium. If there really is an equilibrium, it seems plausible to assume that the y_t variables move together and that z_t is stable. This setup, however, does not exclude the possibility that the y_t variables wander extensively as a group. Thus, they may be driven by a *common stochastic trend*. In other words, it is not excluded that each variable is integrated, yet there exists a linear combination of the variables which is stationary. Integrated variables with this property are called *cointegrated*. In Figure 6.3, two artificially generated cointegrated time series are depicted.

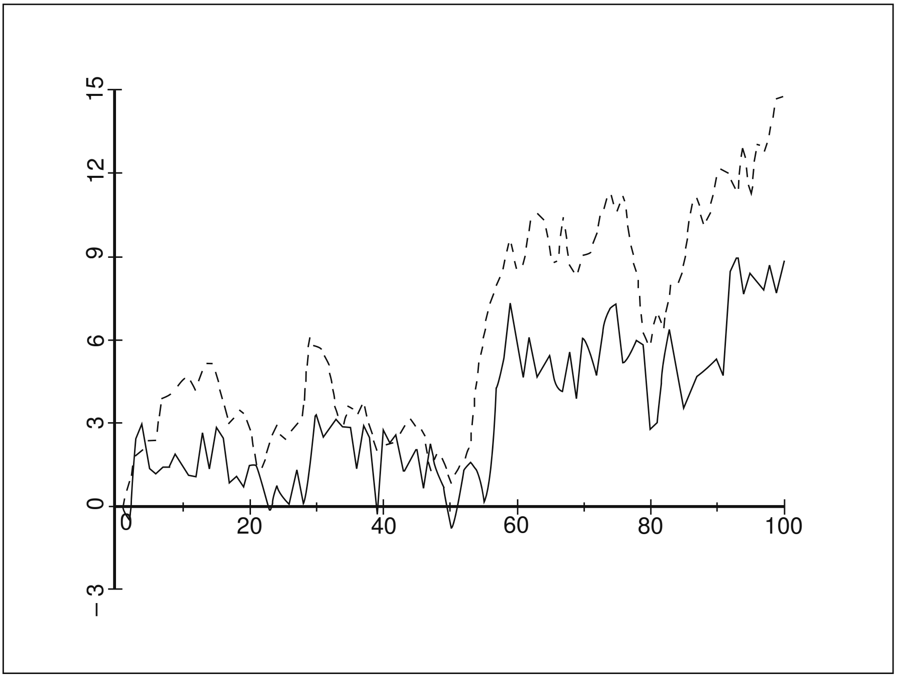

Fig. 6.3. A bivariate cointegrated time series.

Generally, the variables in a K-dimensional process y_t are called *cointegrated of order* (d, b), briefly, $y_t \sim CI(d, b)$, if all components of y_t are $I(d)$ and there exists a linear combination $z_t := \beta' y_t$ with $\beta = (\beta_1, \dots, \beta_K)' \neq 0$ such that z_t is $I(d-b)$. For instance, if all components of y_t are $I(1)$ and $\beta' y_t$ is stationary ($I(0)$), then $y_t \sim CI(1, 1)$. The vector β is called a *cointegrating vector* or a *cointegration vector*. A process consisting of cointegrated variables is called a *cointegrated process*. These processes were introduced by Granger (1981) and Engle & Granger (1987). Since then they have become popular in theoretical and applied econometric work.

In the following, a slightly different definition of cointegration will be used in order to simplify the terminology. We call a K-dimensional process y_t integrated of order d, briefly, $y_t \sim I(d)$, if $\Delta^d y_t$ is stable and $\Delta^{d-1} y_t$ is not stable. The $I(d)$ process y_t is called cointegrated if there is a linear combination $\beta' y_t$ with $\beta \neq 0$ which is integrated of order less than d. This definition differs from the one given by Engle & Granger (1987) in that we do not exclude components of y_t with order of integration less than d. If there is just one $I(d)$ component in y_t and all other components are stable ($I(0)$), then the vector y_t is $I(d)$ according to our definition because $\Delta^d y_t$ is stable and $\Delta^{d-1} y_t$ is not. In such a case a relation $\beta' y_t$ that involves the stationary components only is a cointegration relation in our terms. Clearly, this aspect of our definition is not in line with the original idea of cointegration as a special relation between integrated variables with common stochastic trends. In the following, our definition is still useful because it simplifies the terminology as it avoids distinguishing between variables with different orders of integration. The reader should keep in mind the basic ideas of cointegration when it comes to interpreting specific relationships, however.

Obviously, a cointegrating vector is not unique. Multiplying by a nonzero constant yields a further cointegrating vector. Also, there may be various linearly independent cointegrating vectors. For instance, if there are four variables in a system, the first two may be connected by a long-run equilibrium relation and also the last two. Thus, there may be a cointegrating vector with zeros in the last two positions and one with zeros in the first two positions. In addition, there may be a cointegration relation involving all four variables.

Before the concept of cointegration was introduced, the closely related *error correction models* were discussed in the econometrics literature (see, e.g., Davidson, Hendry, Srba & Yeo (1978), Hendry & von Ungern-Sternberg (1981), Salmon (1982)). In an error correction model, the changes in a variable depend on the deviations from some equilibrium relation. Suppose, for instance, that y_{1t} represents the price of a commodity in a particular market and y_{2t} is the corresponding price of the same commodity in another market. Assume furthermore that the equilibrium relation between the two variables is given by $y_{1t} = \beta_1 y_{2t}$ and that the changes in y_{1t} depend on the deviations from this equilibrium in period $t-1$,

$$\Delta y_{1t} = \alpha_1 (y_{1,t-1} - \beta_1 y_{2,t-1}) + u_{1t}.$$

A similar relation may hold for y_{2t},

$$\Delta y_{2t} = \alpha_2 (y_{1,t-1} - \beta_1 y_{2,t-1}) + u_{2t}.$$

In a more general error correction model, the Δy_{it} may in addition depend on previous changes in both variables as, for instance, in the following model:

$$\begin{aligned} \Delta y_{1t} &= \alpha_1 (y_{1,t-1} - \beta_1 y_{2,t-1}) + \gamma_{11,1} \Delta y_{1,t-1} + \gamma_{12,1} \Delta y_{2,t-1} + u_{1t}, \\ \Delta y_{2t} &= \alpha_2 (y_{1,t-1} - \beta_1 y_{2,t-1}) + \gamma_{21,1} \Delta y_{1,t-1} + \gamma_{22,1} \Delta y_{2,t-1} + u_{2t}. \end{aligned} \tag{6.3.1}$$

Further lags of the Δy_{it}'s may also be included.

To see the close relationship between error correction models and the concept of cointegration, suppose that y_{1t} and y_{2t} are both $I(1)$ variables. In that case all terms in (6.3.1) involving the Δy_{it} are stable. In addition, u_{1t} and u_{2t} are white noise errors which are also stable. Because an unstable term cannot equal a stable process,

$$\alpha_i(y_{1,t-1} - \beta_1 y_{2,t-1}) = \Delta y_{it} - \gamma_{i1,1}\Delta y_{1,t-1} - \gamma_{i2,1}\Delta y_{2,t-1} - u_{it}$$

must be stable too. Hence, if $\alpha_1 \neq 0$ or $\alpha_2 \neq 0$, $y_{1t} - \beta_1 y_{2t}$ is stable and, thus, represents a cointegration relation.

In vector and matrix notation the model (6.3.1) can be written as

$$\Delta y_t = \alpha\beta' y_{t-1} + \mathbf{\Gamma}_1 \Delta y_{t-1} + u_t,$$

or

$$y_t - y_{t-1} = \alpha\beta' y_{t-1} + \mathbf{\Gamma}_1(y_{t-1} - y_{t-2}) + u_t, \tag{6.3.2}$$

where $y_t := (y_{1t}, y_{2t})'$, $u_t := (u_{1t}, u_{2t})'$,

$$\alpha := \begin{bmatrix} \alpha_1 \\ \alpha_2 \end{bmatrix}, \quad \beta' := (1, -\beta_1), \quad \text{and} \quad \mathbf{\Gamma}_1 := \begin{bmatrix} \gamma_{11,1} & \gamma_{12,1} \\ \gamma_{21,1} & \gamma_{22,1} \end{bmatrix}.$$

Rearranging terms in (6.3.2) gives the VAR(2) representation

$$y_t = (I_K + \mathbf{\Gamma}_1 + \alpha\beta')y_{t-1} - \mathbf{\Gamma}_1 y_{t-2} + u_t.$$

Hence, cointegrated variables may be generated by a VAR process.

To see how cointegration can arise more generally in K-dimensional VAR models, consider the VAR(2) process

$$y_t = A_1 y_{t-1} + A_2 y_{t-2} + u_t \tag{6.3.3}$$

with $y_t = (y_{1t}, \dots, y_{Kt})'$. Suppose the process is unstable with

$$|I_K - A_1 z - A_2 z^2| = (1 - \lambda_1 z) \cdots (1 - \lambda_n z) = 0 \quad \text{for } z = 1.$$

Because the λ_i are the reciprocals of the roots of the determinantal polynomial, one or more of them must be equal to 1. All other roots are assumed to lie outside the unit circle, that is, all λ_i which are not 1 are inside the complex unit circle. Because $|I_K - A_1 - A_2| = 0$, the matrix

$$\mathbf{\Pi} := -(I_K - A_1 - A_2)$$

is singular. Suppose $\text{rk}(\mathbf{\Pi}) = r < K$. Then $\mathbf{\Pi}$ can be decomposed as $\mathbf{\Pi} = \alpha\beta'$, where α and β are $(K \times r)$ matrices. From the discussion in the previous section, we know that each variable becomes stationary upon differencing. Let us assume that differencing once is sufficient, subtract y_{t-1} on both sides of (6.3.3) and rearrange terms as

$$y_t - y_{t-1} = -(I_K - A_1 - A_2)y_{t-1} - A_2 y_{t-1} + A_2 y_{t-2} + u_t$$

or

$$\Delta y_t = \mathbf{\Pi} y_{t-1} + \mathbf{\Gamma}_1 \Delta y_{t-1} + u_t, \tag{6.3.4}$$

where $\mathbf{\Gamma}_1 := -A_2$, or

$$\alpha\beta' y_{t-1} = \Delta y_t - \mathbf{\Gamma}_1 \Delta y_{t-1} - u_t.$$

Because the right-hand side involves stationary terms only, $\alpha\beta' y_{t-1}$ must also be stationary and it remains stationary upon multiplication by $(\alpha'\alpha)^{-1}\alpha'$. In other words, $\beta' y_t$ is stationary and, hence, each element of $\beta' y_t$ represents a cointegrating relation. Note that simply taking first differences of all variables in (6.3.3) eliminates the cointegration term which may well contain relations of great importance for a particular analysis. Moreover, in general, a VAR process with cointegrated variables does not admit a pure VAR representation in first differences.

It may also be worth emphasizing that here we have worked under the assumption that all variables are stationary after differencing once. In general, variables with higher integration orders may also be present. In that case, $\beta' y_t$ may not be stationary even if $\text{rk}(\mathbf{\Pi}) = r < K$. The components of y_t may still be cointegrated of a higher order if linear combinations exist which have a reduced order of integration.

In the following, we will be interested in the specific case where all individual variables are $I(1)$ or $I(0)$. The K-dimensional VAR(p) process

$$y_t = A_1 y_{t-1} + \cdots + A_p y_{t-p} + u_t, \tag{6.3.5}$$

is called *cointegrated of rank* r if

$$\mathbf{\Pi} := -(I_K - A_1 - \cdots - A_p)$$

has rank r and, thus, $\mathbf{\Pi}$ can be written as a matrix product $\alpha\beta'$ with α and β being of dimension $(K \times r)$ and of rank r. The matrix β is called a *cointegrating* or *cointegration matrix* or a *matrix of cointegrating* or *cointegration vectors* and α is sometimes called the *loading matrix*. If $r = 0$, Δy_t has a stable VAR($p-1$) representation and, for $r = K$, $|I_K - A_1 - \cdots - A_p| = |-\mathbf{\Pi}| \neq 0$ and, hence, the VAR operator has no unit roots so that y_t is a stable VAR(p) process.

Rewriting (6.3.5) as in (6.3.4) it has a *vector error correction model* (*VECM*) representation

$$\begin{aligned} \Delta y_t &= \mathbf{\Pi} y_{t-1} + \mathbf{\Gamma}_1 \Delta y_{t-1} + \cdots + \mathbf{\Gamma}_{p-1} \Delta y_{t-p+1} + u_t \\ &= \alpha\beta' y_{t-1} + \mathbf{\Gamma}_1 \Delta y_{t-1} + \cdots + \mathbf{\Gamma}_{p-1} \Delta y_{t-p+1} + u_t, \end{aligned} \tag{6.3.6}$$

where

$$\mathbf{\Gamma}_i := -(A_{i+1} + \cdots + A_p), \quad i = 1, \ldots, p-1.$$

If this representation of a cointegrated process is given, it is easy to recover the corresponding VAR form (6.3.5) by noting that

$$\begin{aligned} A_1 &= \mathbf{\Pi} + I_K + \mathbf{\Gamma}_1 \\ A_i &= \mathbf{\Gamma}_i - \mathbf{\Gamma}_{i-1}, \quad i = 2, \ldots, p-1, \\ A_p &= -\mathbf{\Gamma}_{p-1}. \end{aligned} \tag{6.3.7}$$

It may be worth pointing out that we can also rearrange the terms in a different way and obtain a representation

$$\Delta y_t = D_1 \Delta y_{t-1} + \cdots + D_{p-1} \Delta y_{t-p+1} + \mathbf{\Pi} y_{t-p} + u_t, \tag{6.3.8}$$

where the error correction term appears at lag p and

$$D_i = -(I_K - A_1 - \cdots - A_i), \quad i = 1, \ldots, p-1.$$

In the following sections, we will usually work with (6.3.5) or (6.3.6). Of course, thereby we work within a much more narrow framework than that allowed for in the general definition of cointegration. First, we consider $I(1)$ processes only and, second, the discussion is limited to finite order VAR processes or VECMs.

It is important to note that the decomposition of the $(K \times K)$ matrix $\mathbf{\Pi}$ as the product of two $(K \times r)$ matrices, $\mathbf{\Pi} = \alpha\beta'$, is not unique. In fact, for every nonsingular $(r \times r)$ matrix Q, we can define $\alpha^* = \alpha Q'$ and $\beta^* = \beta Q^{-1}$ and get $\mathbf{\Pi} = \alpha^*\beta^{*\prime}$. This nonuniqueness of the decomposition of $\mathbf{\Pi}$ shows again that the cointegration relations are not unique. It is possible, however, to impose restrictions on β and/or α to get unique relations. Such restrictions may be implied by subject matter considerations or they may be imposed for convenience, using the algebraic properties of the associated matrices.

As an example, consider a system of three interest rates, $y_t = (y_{1t}, y_{2t}, y_{3t})'$, where y_{1t} is a short-term rate, y_{2t} is a medium-term rate, and y_{3t} is a long-term rate. Suppose all three interest rates are $I(1)$ variables whereas the interest rate spreads, $y_{it} - y_{jt}$ $(i \neq j)$ are stationary $(I(0))$. Then we have two linearly independent cointegrating relations which can, for example, be written as

$$\beta' y_t = \begin{bmatrix} 1 & 0 & -1 \\ 0 & 1 & -1 \end{bmatrix} y_t$$

or, alternatively, as

$$\beta^{*\prime} y_t = \begin{bmatrix} 1 & -1 & 0 \\ 0 & 1 & -1 \end{bmatrix} y_t.$$

Using the fact that $\text{rk}(\beta) = r$, there must be r linearly independent rows. Thus, by a suitable rearrangement of the variables it can always be ensured that the first r rows of β are linearly independent. Hence, the upper $(r \times r)$

submatrix consisting of the first r rows of β is nonsingular. Choosing Q then equal to this matrix gives a cointegration matrix

$$\beta^* = \begin{bmatrix} I_r \\ \beta_{(K-r)} \end{bmatrix}, \tag{6.3.9}$$

where $\beta_{(K-r)}$ is $((K-r)\times r)$. This normalization will occasionally be used in the following because it is quite convenient to ensure a unique cointegration matrix. It does not imply a loss of generality except that it is assumed that the variables are arranged in the right way so that the normalization is feasible. If the system is known, as implicitly assumed here, rearranging the variables in a suitable way is no problem, of course. In fact, we just need to know the cointegration properties between all subsets of variables in order to find a suitable arrangement of the variables.

To see this, consider again a three-dimensional system, $y_t = (y_{1t}, y_{2t}, y_{3t})'$, with cointegrating rank 1 so that there is just one cointegration vector β. In that case, the normalization in (6.3.9) amounts to setting the first component of the cointegration vector to one. Hence, $\beta^{*\prime} y_t = [1, \beta'_{(K-1)}]y_t = y_{1t} + \beta_2 y_{2t} + \beta_3 y_{3t}$. Clearly, this normalization is only feasible if the first component of y_t actually belongs to the cointegration relation and has nonzero coefficient. If we know that y_{2t} and y_{3t} are not cointegrated while y_{1t}, y_{2t}, and y_{3t} together are cointegrated, then we know already that y_{1t} is part of the cointegration relation and, thus, has a nonzero coefficient in β.

As another example, suppose y_t has cointegrating rank 2. In that case the normalized cointegrating relations are

$$\begin{bmatrix} 1 & 0 & \beta_1 \\ 0 & 1 & \beta_2 \end{bmatrix} y_t = \begin{bmatrix} y_{1t} + \beta_1 y_{3t} \\ y_{2t} + \beta_2 y_{3t} \end{bmatrix}.$$

Thus, a cointegration relation must exist in the bivariate systems $(y_{1t}, y_{3t})'$ and $(y_{2t}, y_{3t})'$. By checking these subsystems separately, a possible ordering of the variables is easy to find. It may be worth mentioning, however, that given our general definition of cointegration, it is possible that in this example y_{1t} or/and y_{2t} are in fact stationary $I(0)$ variables. For instance, if both are $I(0)$, $\beta_1 = \beta_2 = 0$. Recall that a process y_t is called $I(1)$ even if only a single component is $I(1)$ and the other components are $I(0)$.

Generally, any stationary variables in the system must be placed in the upper r-dimensional subvector of y_t. If y_{kt}, the k-th component of y_t, is stationary, there is a 'cointegrating relation' $\beta'_k y_t$ with β_k being a vector with a one as the k-th component and zeros elsewhere so that $\beta'_k y_t = y_{kt}$. Thus, there is a cointegrating relation for each of the stationary components of y_t. Because the associated cointegrating vectors are linearly independent, the cointegrating rank must be at least as great as the number of $I(0)$ variables in the system.

The important result to remember from this discussion is that the normalization of the cointegration matrix given in (6.3.9) is always possible if the variables are arranged in a suitable way. Finding the proper ordering is easy if the cointegration properties of all subsystems are known, including the univariate subsystems. In other words, we also need to know the order of integration of the individual variables in the system. In practice, the order of integration and the cointegrating rank of a given system and its subsystems will not be known. Statistical procedures for determining the cointegrating rank which can help to overcome this practical problem are discussed in Chapter 8.

If the normalization in (6.3.9) is made, the system may also be set up as

$$\begin{aligned} y_t^{(1)} &= -\beta'_{(K-r)} y_t^{(2)} + z_t^{(1)}, \\ \Delta y_t^{(2)} &= z_t^{(2)}, \end{aligned} \tag{6.3.10}$$

where $y_t^{(1)}$ and $z_t^{(1)}$ are $(r \times 1)$, $y_t^{(2)}$ and $z_t^{(2)}$ are $((K-r) \times 1)$ and $z_t = (z_t^{(1)\prime}, z_t^{(2)\prime})'$ is a stationary process. There cannot be any cointegrating relations between the components of the subsystem $y_t^{(2)}$, because otherwise there would be more than r linearly independent cointegrating relations and the cointegrating rank would be larger than r. Thus, the variables in $y_t^{(2)}$ represent stochastic trends in the system. The representation (6.3.10) is known as the *triangular representation* of a cointegrated system. It has been used extensively in some of the literature related to cointegration analysis (see, e.g., Park & Phillips (1988, 1989)).

Yet another useful representation of a cointegrated system is given by Johansen (1995, Theorem 4.2). The underlying result is often referred to as *Granger representation theorem.* To state this representation, we use the following notation. For $m \geq n$, we denote by $M_\perp$ an orthogonal complement of the $(m \times n)$ matrix M with $\text{rk}(M) = n$ (see also Appendix A.8.2). In other words, $M_\perp$ is any $(m \times (m-n))$ matrix with $\text{rk}(M_\perp) = m-n$ and $M'M_\perp = 0$. If M is a nonsingular square matrix $(m = n)$, then $M_\perp = 0$ and if $n = 0$, we define $M_\perp = I_m$. This latter convention is sometimes useful to avoid clumsy notation and looking at different cases separately. We assume that y_t is a K-dimensional cointegrated $I(1)$ process as in (6.3.6) with cointegration rank r, $0 \leq r < K$. Then the following proposition holds.

Proposition 6.1 (*Granger Representation Theorem*)
Suppose

$$\Delta y_t = \alpha\beta' y_{t-1} + \mathbf{\Gamma}_1 \Delta y_{t-1} + \cdots + \mathbf{\Gamma}_{p-1} \Delta y_{t-p+1} + u_t, \quad t = 1, 2, \ldots,$$

where $y_t = 0$ for $t \leq 0$, u_t is white noise for $t = 1, 2, \ldots$, and $u_t = 0$ for $t \leq 0$. Moreover, define

$$C(z) := (1-z)I_K - \alpha\beta' z - \sum_{i=1}^{p-1} \mathbf{\Gamma}_i (1-z) z^i$$

and let the following conditions hold for the parameters:

(a) $\det C(z) = 0 \Rightarrow |z| > 1$ or $z = 1$.
(b) The number of unit roots, $z = 1$, is exactly $K - r$.
(c) α and β are $(K \times r)$ matrices with $\text{rk}(\alpha) = \text{rk}(\beta) = r$.

Then y_t has the representation

$$y_t = \Xi \sum_{i=1}^{t} u_i + \Xi^*(L)u_t + y_0^*, \tag{6.3.11}$$

where

$$\Xi = \beta_\perp \left[\alpha'_\perp \left(I_K - \sum_{i=1}^{p-1} \mathbf{\Gamma}_i \right) \beta_\perp \right]^{-1} \alpha'_\perp, \tag{6.3.12}$$

$\Xi^*(L)u_t = \sum_{j=0}^{\infty} \Xi_j^* u_{t-j}$ is an $I(0)$ process and y_0^* contains initial values. ■

Remark 1 The proposition is of fundamental importance because it decomposes the process y_t into $I(1)$ and $I(0)$ components which have to be treated accordingly, for example, when asymptotic properties of parameter estimators are derived (see Chapter 7). It makes precise under what conditions the process y_t is driven by $K - r$ $I(1)$ components and r $I(0)$ components. The representation in (6.3.11) is a multivariate version of the *Beveridge-Nelson decomposition* of y_t. The first term on the right-hand side of (6.3.11) consists of K random walks $\sum_{i=1}^{t} u_i$ which are multiplied by a matrix of rank $K - r$, denoted by Ξ. Thus, there are actually $K - r$ stochastic trends driving the system. They determine to a large extent the development of y_t. Therefore one may call y_t an $I(1)$ process if there are actually $I(1)$ trends (random walks) in the representation (6.3.11). In other words, y_t is $I(1)$ if it has the representation (6.3.11) with $\Xi \neq 0$. Clearly, for Ξ to have the form given in (6.3.12), the $((K - r) \times (K - r))$ matrix

$$\alpha'_\perp \left(I_K - \sum_{i=1}^{p-1} \mathbf{\Gamma}_i \right) \beta_\perp$$

must be invertible. Only under that condition, $\text{rk}(\Xi) = K - r$. Therefore the latter condition ensures that y_t is actually driven by $K - r$ random walk components. ■

Remark 2 The parameter matrices Ξ_j^* in (6.3.11) are determined by the model parameters. To state the precise relation, we define

$$\bar{\beta} := \beta(\beta'\beta)^{-1} \quad (K \times r),$$

$$\underset{(K \times K)}{Q} := \begin{bmatrix} \beta' \\ \beta'_\perp \end{bmatrix} \quad \text{so that} \quad Q^{-1} = [\bar{\beta} : \beta_\perp],$$

$$\begin{aligned}
\mathbf{\Gamma}(z) &:= I_K - \sum_{i=1}^{p-1} \mathbf{\Gamma}_i z^i, \\
B_*(z) &:= Q[\mathbf{\Gamma}(z)\bar{\beta}(1-z) - \alpha z : \mathbf{\Gamma}(z)\beta_\perp], \\
B(z) &= I_K - \sum_{i=1}^{p} B_i z^i := Q^{-1} B_*(z) Q,
\end{aligned} \tag{6.3.13}$$

and

$$\mathbf{\Theta}(z) := B(z)^{-1} = \sum_{j=0}^{\infty} \mathbf{\Theta}_j z^j.$$

Notice that $B(0) = Q^{-1}B_*(0)Q = [\bar{\beta} : \beta_\perp]Q = I_K$. Hence, $B(z)$ has the representation $I_K - \sum_{i=1}^{p} B_i z^i$ stated in (6.3.13). Moreover, the matrix operator $\mathbf{\Theta}(z)$ can be decomposed as

$$\mathbf{\Theta}(z) = \mathbf{\Theta}(1) + (1-z)\mathbf{\Theta}^*(z),$$

where expressions for the $\mathbf{\Theta}_j^*$'s can be found by comparing coefficients in $\mathbf{\Theta}(z) = \sum_{j=0}^{\infty} \mathbf{\Theta}_j z^j$ and

$$\begin{aligned}
\mathbf{\Theta}(1) + (1-z)\mathbf{\Theta}^*(z) &= \mathbf{\Theta}(1) + \sum_{j=0}^{\infty} \mathbf{\Theta}_j^* z^j (1-z) \\
&= (\mathbf{\Theta}(1) + \mathbf{\Theta}_0^*) + \sum_{j=1}^{\infty} (\mathbf{\Theta}_j^* - \mathbf{\Theta}_{j-1}^*) z^j.
\end{aligned}$$

Hence,

$$\mathbf{\Theta}_0 = \mathbf{\Theta}(1) + \mathbf{\Theta}_0^*$$

and

$$\mathbf{\Theta}_i = \mathbf{\Theta}_i^* - \mathbf{\Theta}_{i-1}^*, \quad i = 1, 2, \dots.$$

Using the last expression, we get by successive substitution,

$$\begin{aligned}
\mathbf{\Theta}_i^* &= \mathbf{\Theta}_i + \mathbf{\Theta}_{i-1}^* = \sum_{j=1}^{i} \mathbf{\Theta}_{i-j} + \mathbf{\Theta}_0^* \\
&= \sum_{j=1}^{i} \mathbf{\Theta}_{i-j} + \mathbf{\Theta}_0 - \mathbf{\Theta}(1) = -\sum_{j=i+1}^{\infty} \mathbf{\Theta}_j, \quad i = 1, 2, \dots.
\end{aligned} \tag{6.3.14}$$

From these quantities the operator $\mathbf{\Xi}^*(z)$ in (6.3.11) can be obtained as

$$\mathbf{\Xi}^*(z) = [\mathbf{\Theta}^*(z) + \bar{\beta}\beta' B(z)^{-1}] \tag{6.3.15}$$

(see the proof of Proposition 6.1). The representation (6.3.11) will turn out to be useful, for example, in Chapter 9, where structural VECMs are discussed. The coefficient matrices $\mathbf{\Xi}_j^*$ of the operator $\mathbf{\Xi}^*(z)$ will then play an important role as specific impulse response coefficients. ■

Proof of Proposition 6.1
The proof is adapted from Saikkonen (2005). We use the notation from Remark 2 and first show that under the conditions of Proposition 6.1,

$$C(z) = Q^{-1}B_*(z)P(z), \tag{6.3.16}$$

where

$$P(z) := \begin{bmatrix} \beta' \\ (1-z)\bar{\beta}'_\perp \end{bmatrix} = \begin{bmatrix} I_r & 0 \\ 0 & (1-z)I_{K-r} \end{bmatrix} Q.$$

This representation is obtained by noting that

$$\begin{aligned} C(z) &= [\mathbf{\Gamma}(z)(1-z) - \alpha\beta' z]Q^{-1}Q \\ &= [\mathbf{\Gamma}(z)\bar{\beta}(1-z) - \alpha\beta'\bar{\beta}z : \mathbf{\Gamma}(z)\beta_\perp(1-z) - \alpha\beta'\beta_\perp z]Q \\ &= [\mathbf{\Gamma}(z)\bar{\beta}(1-z) - \alpha z : \mathbf{\Gamma}(z)\beta_\perp(1-z)] \begin{bmatrix} \beta' \\ \bar{\beta}'_\perp \end{bmatrix} \\ &= Q^{-1}Q[\mathbf{\Gamma}(z)\bar{\beta}(1-z) - \alpha z : \mathbf{\Gamma}(z)\beta_\perp] \begin{bmatrix} \beta' \\ (1-z)\bar{\beta}'_\perp \end{bmatrix}. \end{aligned}$$

Clearly, $\det P(z)$ has exactly $K-r$ unit roots and, thus, $\det B_*(z)$ cannot have any such roots so that $\det B_*(z) \neq 0$ for $|z| \leq 1$ must hold. In other words, $B_*(L)$ is an invertible operator.

Now define

$$z_t := Q^{-1}P(L)y_t = \bar{\beta}\beta' y_t + \beta_\perp\bar{\beta}'_\perp \Delta y_t \tag{6.3.17}$$

and note that

$$\beta' z_t = \beta' y_t. \tag{6.3.18}$$

For the operator $B(z) = Q^{-1}B_*(z)Q$, we have $B(0) = Q^{-1}B_*(0)Q = I_K$ and $\det B(z) \neq 0$ for $|z| \leq 1$ because $\det B_*(z)$ has no roots inside or on the complex unit circle. Moreover,

$$B(L)z_t = Q^{-1}B_*(L)QQ^{-1}P(L)y_t = C(L)y_t = u_t.$$

Thus,

$$z_t = \sum_{i=1}^{p} B_i z_{t-i} + u_t \tag{6.3.19}$$

is a stable VAR(p) process with the same residual process u_t as y_t. We know from Chapter 2 that it has an MA representation

$$z_t = B(L)^{-1}u_t = \mathbf{\Theta}(L)u_t = \sum_{j=0}^{\infty} \mathbf{\Theta}_j u_{t-j}. \tag{6.3.20}$$

As we have seen in Remark 2, the matrix operator $\mathbf{\Theta}(z)$ can be decomposed as

$$\mathbf{\Theta}(z) = \mathbf{\Theta}(1) + (1-z)\mathbf{\Theta}^*(z).$$

Hence, we get from (6.3.20),

$$z_t = \mathbf{\Theta}(1)u_t + \mathbf{\Theta}^*(L)\Delta u_t = B(1)^{-1}u_t + \mathbf{\Theta}^*(L)\Delta u_t. \tag{6.3.21}$$

Using

$$y_t = Q^{-1}Qy_t = [\bar{\beta} : \beta_\perp] \begin{bmatrix} \beta' y_t \\ \bar{\beta}'_\perp y_t \end{bmatrix} = \bar{\beta}\beta' y_t + \beta_\perp \bar{\beta}'_\perp y_t$$

and, hence,

$$\Delta y_t = \bar{\beta}\beta'\Delta y_t + \beta_\perp\bar{\beta}'_\perp \Delta y_t,$$

it follows from (6.3.17) and (6.3.18) that $\Delta y_t = z_t - \bar{\beta}\beta' z_{t-1}$. Thus,

$$\beta_\perp\bar{\beta}'_\perp \Delta y_t = \beta_\perp\bar{\beta}'_\perp z_t.$$

Substituting the expression from (6.3.21) for z_t gives

$$\begin{aligned} \Delta y_t &= \beta_\perp\bar{\beta}'_\perp z_t + \bar{\beta}\beta'\Delta y_t \\ &= \beta_\perp\bar{\beta}'_\perp B(1)^{-1}u_t + \mathbf{\Theta}^*(L)\Delta u_t + \bar{\beta}\beta'\Delta y_t := w_t. \end{aligned}$$

Solving for $y_t = y_{t-1} + w_t$ results in

$$\begin{aligned} y_t &= y_0 + \sum_{i=1}^t w_i \\ &= y_0 + \beta_\perp\bar{\beta}'_\perp B(1)^{-1}\sum_{i=1}^t u_i + \mathbf{\Theta}^*(L)\sum_{i=1}^t \Delta u_t + \bar{\beta}\beta'\sum_{i=1}^t \Delta y_t \\ &= y_0 + \beta_\perp\bar{\beta}'_\perp B(1)^{-1}\sum_{i=1}^t u_i + \mathbf{\Theta}^*(L)(u_t - u_0) + \bar{\beta}\beta'(y_t - y_0) \\ &= \beta_\perp\bar{\beta}'_\perp B(1)^{-1}\sum_{i=1}^t u_i + \mathbf{\Theta}^*(L)u_t + \bar{\beta}\beta' y_t + y_0^*, \end{aligned} \tag{6.3.22}$$

where $y_0^* := y_0 - \mathbf{\Theta}^*(L)u_0 - \bar{\beta}\beta' y_0$. Using $\beta' y_t = \beta' z_t$, the term $\bar{\beta}\beta' y_t = \bar{\beta}\beta' z_t$ is seen to have a representation

$$\bar{\beta}\beta' z_t = \bar{\beta}\beta'\mathbf{\Theta}(L)u_t$$

and, thus, $\mathbf{\Theta}^*(L)u_t + \bar{\beta}\beta' y_t$ has an MA representation

$$\mathbf{\Xi}^*(L)u_t = [\mathbf{\Theta}^*(L) + \bar{\beta}\beta'\mathbf{\Theta}(L)]u_t.$$

For the first term on the right-hand side of (6.3.22) we have

$$\begin{aligned}
\beta_\perp \bar{\beta}_\perp' B(1)^{-1} &= \beta_\perp \bar{\beta}_\perp' Q^{-1} B_*(1)^{-1} Q \\
&= \beta_\perp \bar{\beta}_\perp' [\bar{\beta} : \beta_\perp][-\alpha : \mathbf{\Gamma}(1)\beta_\perp]^{-1} \\
&= \beta_\perp [0 : I_{K-r}][-\alpha : \mathbf{\Gamma}(1)\beta_\perp]^{-1} \\
&= \beta_\perp [\alpha_\perp' \mathbf{\Gamma}(1)\beta_\perp]^{-1} \alpha_\perp',
\end{aligned}$$

because

$$[-\alpha : \mathbf{\Gamma}(1)\beta_\perp]^{-1} = \begin{bmatrix} (\alpha'\alpha)^{-1}\alpha'\{\mathbf{\Gamma}(1)\beta_\perp[\alpha_\perp'\mathbf{\Gamma}(1)\beta_\perp]^{-1}\alpha_\perp' - I_K\} \\ [\alpha_\perp'\mathbf{\Gamma}(1)\beta_\perp]^{-1}\alpha_\perp' \end{bmatrix}.$$

Hence, $\Xi = \beta_\perp \bar{\beta}_\perp' B(1)^{-1}$ is as stated in the proposition. Notice that the invertibility of $\alpha_\perp' \mathbf{\Gamma}(1)\beta_\perp$ follows from the invertibility of $B(1)$ which in turn is implied by $\det B(z) \neq 0$ for $|z| \leq 1$. ■

6.4 Deterministic Terms in Cointegrated Processes

In the previous section, we have ignored deterministic terms in the DGP. Clearly, deterministic terms may also be present in cointegrated processes and VECMs. Actually, from the discussion of the random walk with drift it should be clear that deterministic terms in a VAR process with unit roots may have a different impact than in a stable VAR. For example, an intercept term in a random walk generates a linear trend in the mean of the process, whereas an intercept term in a stable AR process just implies a constant mean value. To explore the implications of the deterministic term, the following model is assumed:

$$y_t = \mu_t + x_t, \tag{6.4.1}$$

where x_t is a zero mean VAR(p) process with possibly cointegrated variables and μ_t stands for the deterministic term. For example, the deterministic term may just be a constant, $\mu_t = \mu_0$, or it may be a linear trend term, $\mu_t = \mu_0 + \mu_1 t$, where μ_0 and μ_1 are fixed K-dimensional parameter vectors. Other possible deterministic terms that may be included are seasonal dummy variables or other dummies to account for special events. The advantage of setting up the process in the form (6.4.1) by adding the deterministic part to the zero mean stochastic part is that the mean of the y_t variables is clearly specified by the deterministic term and does not need to be derived from quantities that involve the parameters of the stochastic part in addition. The disadvantage is that the stochastic part x_t is not directly observable in general. Therefore, for estimation purposes, for instance, we have to rewrite the process in terms of the observable y_t's. We will do so in the following for some cases of specific interest.

It is assumed that the DGP of x_t can be represented as a VECM such as (6.3.6),

$$\begin{aligned} \Delta x_t &= \alpha\beta' x_{t-1} + \boldsymbol{\Gamma}_1 \Delta x_{t-1} + \cdots + \boldsymbol{\Gamma}_{p-1} \Delta x_{t-p+1} + u_t \\ &= \boldsymbol{\Pi} x_{t-1} + \boldsymbol{\Gamma}_1 \Delta x_{t-1} + \cdots + \boldsymbol{\Gamma}_{p-1} \Delta x_{t-p+1} + u_t. \end{aligned} \tag{6.4.2}$$

Considering now the case of a constant deterministic term, $\mu_t = \mu_0$, we have $x_t = y_t - \mu_0$ so that $\Delta y_t = \Delta x_t$ and from (6.4.2) we get

$$\begin{aligned} \Delta y_t &= \alpha\beta'(y_{t-1} - \mu_0) + \boldsymbol{\Gamma}_1 \Delta y_{t-1} + \cdots + \boldsymbol{\Gamma}_{p-1} \Delta y_{t-p+1} + u_t \\ &= \alpha\beta^{o\prime} \begin{bmatrix} y_{t-1} \\ 1 \end{bmatrix} + \boldsymbol{\Gamma}_1 \Delta y_{t-1} + \cdots + \boldsymbol{\Gamma}_{p-1} \Delta y_{t-p+1} + u_t \\ &= \boldsymbol{\Pi}^o y^o_{t-1} + \boldsymbol{\Gamma}_1 \Delta y_{t-1} + \cdots + \boldsymbol{\Gamma}_{p-1} \Delta y_{t-p+1} + u_t, \end{aligned} \tag{6.4.3}$$

where $\beta^{o\prime} := [\beta' : \tau']$ with $\tau' := -\beta'\mu_0$ an $(r \times 1)$ vector,

$$y^o_{t-1} := \begin{bmatrix} y_{t-1} \\ 1 \end{bmatrix}$$

and $\boldsymbol{\Pi}^o := [\boldsymbol{\Pi} : \nu_0]$ is $(K \times (K+1))$ with $\nu_0 := -\boldsymbol{\Pi}\mu_0 = \alpha\tau'$. Hence, if there is just a constant mean, it can be absorbed into the cointegration relations. In other words, the constant mean becomes an intercept term in the cointegration relations. Of course, the model can also be written with an overall intercept term as

$$\begin{aligned} \Delta y_t &= \nu_0 + \alpha\beta' y_{t-1} + \boldsymbol{\Gamma}_1 \Delta y_{t-1} + \cdots + \boldsymbol{\Gamma}_{p-1} \Delta y_{t-p+1} + u_t \\ &= \nu_0 + \boldsymbol{\Pi} y_{t-1} + \boldsymbol{\Gamma}_1 \Delta y_{t-1} + \cdots + \boldsymbol{\Gamma}_{p-1} \Delta y_{t-p+1} + u_t. \end{aligned} \tag{6.4.4}$$

Here ν_0 cannot be an arbitrary $(K \times 1)$ vector but has to satisfy the indicated restrictions ($\nu_0 = \alpha\tau'$) in order to ensure that the intercept term in this model does not generate a linear trend in the mean of the y_t variables. By specifying the deterministic term in additive form as in (6.4.1), the properties of the mean of y_t are easy to see.

A process with a linear trend in the mean, $\mu_t = \mu_0 + \mu_1 t$, is another case of practical importance. Using $x_t = y_t - \mu_0 - \mu_1 t$, $\Delta x_t = \Delta y_t - \mu_1$, and (6.4.2), gives

$$\begin{aligned} \Delta y_t - \mu_1 &= \alpha\beta'(y_{t-1} - \mu_0 - \mu_1(t-1)) + \boldsymbol{\Gamma}_1(\Delta y_{t-1} - \mu_1) + \cdots \\ &\quad + \boldsymbol{\Gamma}_{p-1}(\Delta y_{t-p+1} - \mu_1) + u_t \end{aligned} \tag{6.4.5}$$

or, collecting deterministic terms,

$$\begin{aligned} \Delta y_t &= \nu + \alpha[\beta' : \eta'] \begin{bmatrix} y_{t-1} \\ t-1 \end{bmatrix} + \boldsymbol{\Gamma}_1 \Delta y_{t-1} + \cdots + \boldsymbol{\Gamma}_{p-1} \Delta y_{t-p+1} + u_t \\ &= \nu + \boldsymbol{\Pi}^+ y^+_{t-1} + \boldsymbol{\Gamma}_1 \Delta y_{t-1} + \cdots + \boldsymbol{\Gamma}_{p-1} \Delta y_{t-p+1} + u_t, \end{aligned} \tag{6.4.6}$$

where $\nu := -\boldsymbol{\Pi}\mu_0 + (I_K - \boldsymbol{\Gamma}_1 - \cdots - \boldsymbol{\Gamma}_{p-1})\mu_1$, $\eta' := -\beta'\mu_1$, $\boldsymbol{\Pi}^+ := \alpha[\beta' : \eta']$ is a $(K \times (K+1))$ matrix and

$$y_t^+ := \begin{bmatrix} y_t \\ t \end{bmatrix}.$$

Now the general intercept term ν is in fact unrestricted and can take on any value from $\mathbb{R}^K$, depending of course on μ_0, μ_1, and the other parameters. In contrast, the trend term can be absorbed into the cointegration relations. Writing the model with unrestricted linear trend term in the form

$$\Delta y_t = \nu_0 + \nu_1 t + \mathbf{\Pi} y_{t-1} + \mathbf{\Gamma}_1 \Delta y_{t-1} + \cdots + \mathbf{\Gamma}_{p-1} \Delta y_{t-p+1} + u_t,$$

the model is actually in principle capable of generating quadratic trends in the means of the variables.

It is also possible, that the trend slope parameter μ_1 is orthogonal to the cointegration matrix so that $\beta'\mu_1 = 0$ and, hence, $\eta = 0$ and the trend term disappears from the cointegration relations. This situation can also occur if $\mu_1 \neq 0$ and the variables actually have linear trends in their means. The linear trends will then be generated via the intercept term ν. The resulting model,

$$\begin{aligned} \Delta y_t &= \nu + \alpha\beta' y_{t-1} + \mathbf{\Gamma}_1 \Delta y_{t-1} + \cdots + \mathbf{\Gamma}_{p-1} \Delta y_{t-p+1} + u_t \\ &= \nu + \mathbf{\Pi} y_{t-1} + \mathbf{\Gamma}_1 \Delta y_{t-1} + \cdots + \mathbf{\Gamma}_{p-1} \Delta y_{t-p+1} + u_t, \end{aligned} \tag{6.4.7}$$

with unrestricted intercept term ν will be of some importance later on. It represents a situation where a linear trend appears in the variables but not in the cointegration relations. Notice, however, that in this situation the cointegration rank must be smaller than K. If the process has cointegrating rank K, it is stable and, hence, it cannot generate a linear trend when just an intercept is included in the model. Formally, a "cointegrating matrix" β of rank K is nonsingular so that $\beta'\mu_1$ cannot be zero if μ_1 is nonzero.

It may also be worth noting that the specification of the deterministic component in additive form as in (6.4.1) has the additional advantage that the Beveridge-Nelson representation of y_t is obtained by adding the deterministic term to the Beveridge-Nelson representation of x_t. Thus, a suitable generalization of the Granger representation theorem (Proposition 6.1) is readily available.

6.5 Forecasting Integrated and Cointegrated Variables

If forecasting is the objective, the VAR form of a process is quite convenient. Because forecasting the deterministic part is trivial, a purely stochastic process will be considered initially. For a VAR(p) process,

$$y_t = A_1 y_{t-1} + \cdots + A_p y_{t-p} + u_t, \tag{6.5.1}$$

the optimal h-step forecast with minimal MSE is given by the conditional expectation, provided that expectation exists, even if $\det(I_K - A_1 z - \cdots - A_p z^p)$ has roots on the unit circle. In the proof of the optimality of the conditional expectation in Section 2.2.2, we have not used the stationarity and stability of the system. Thus, assuming that u_t is independent white noise, the optimal h-step forecast at origin t is

$$y_t(h) = A_1 y_t(h-1) + \cdots + A_p y_t(h-p), \tag{6.5.2}$$

where $y_t(j) := y_{t+j}$ for $j \leq 0$, just as in the stationary, stable case.

Also the forecast errors are of the same form as in the stable case. To see this, we write the process (6.5.1) in VAR(1) form as

$$Y_t = \mathbf{A} Y_{t-1} + U_t, \tag{6.5.3}$$

where

$$Y_t := \underset{(Kp\times 1)}{\begin{bmatrix} y_t \\ \vdots \\ y_{t-p+1} \end{bmatrix}}, \ \mathbf{A} := \underset{(Kp\times Kp)}{\begin{bmatrix} A_1 & A_2 & \dots & A_{p-1} & A_p \\ I_K & 0 & \dots & 0 & 0 \\ 0 & I_K & & 0 & 0 \\ \vdots & & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & I_K & 0 \end{bmatrix}}, \text{ and } U_t := \underset{(Kp\times 1)}{\begin{bmatrix} u_t \\ 0 \\ \vdots \\ 0 \end{bmatrix}}.$$

If u_t is independent white noise, the optimal h-step forecast of Y_t is

$$Y_t(h) = \mathbf{A} Y_t(h-1) = \mathbf{A}^h Y_t.$$

Moreover,

$$\begin{aligned} Y_{t+h} &= \mathbf{A} Y_{t+h-1} + U_{t+h} \\ &= \mathbf{A}^h Y_t + U_{t+h} + \mathbf{A} U_{t+h-1} + \cdots + \mathbf{A}^{h-1} U_{t+1}. \end{aligned}$$

Hence, the forecast error for the process Y_t is

$$Y_{t+h} - Y_t(h) = U_{t+h} + \mathbf{A} U_{t+h-1} + \cdots + \mathbf{A}^{h-1} U_{t+1}.$$

Premultiplying by the $(K \times Kp)$ matrix $J := [I_K : 0 : \cdots : 0]$ gives

$$\begin{aligned} y_{t+h} - y_t(h) &= JU_{t+h} + J\mathbf{A}J'JU_{t+h-1} + \cdots + J\mathbf{A}^{h-1}J'JU_{t+1} \\ &= u_{t+h} + \Phi_1 u_{t+h-1} + \cdots + \Phi_{h-1} u_{t+1}, \end{aligned} \tag{6.5.4}$$

where $J'JU_t = U_t$ and $\Phi_i = J\mathbf{A}^i J'$ have been used. Thus, the form of the forecast error is exactly the same as in the stable case and the forecast is easily seen to be unbiased, that is,

$$E[y_{t+h} - y_t(h)] = 0.$$

Furthermore, the Φ_i's may be obtained from the A_i's by the recursions

$$\Phi_i = \sum_{j=1}^{i} \Phi_{i-j} A_j, \quad i = 1, 2, \dots, \tag{6.5.5}$$

with $\Phi_0 = I_K$, just as in Chapter 2. Also the forecast MSE matrix becomes

$$\Sigma_y(h) = \sum_{i=0}^{h-1} \Phi_i \Sigma_u \Phi_i', \tag{6.5.6}$$

as in the stable case. Yet there is a very important difference. In the stable case, the Φ_i's converge to zero as $i \to \infty$ and $\Sigma_y(h)$ converges to the covariance matrix of y_t as $h \to \infty$. This result was obtained because the eigenvalues of $\mathbf{A}$ have modulus less than one in the stable case. Hence, $\Phi_i = J\mathbf{A}^i J' \to 0$ as $i \to \infty$. Because the eigenvalues of $\mathbf{A}$ are just the reciprocals of the roots of the determinantal polynomial $\det(I_K - A_1 z - \cdots - A_p z^p)$, the Φ_i's do not converge to zero in the presently considered unstable case where one or more of the eigenvalues of $\mathbf{A}$ are 1. Consequently, some elements of the forecast MSE matrix $\Sigma_y(h)$ will approach infinity as $h \to \infty$. In other words, the forecast MSEs will be unbounded and the forecast uncertainty may become extremely large as we make forecasts for the distant future, even if the structure of the process does not change.

To illustrate this point, consider the following bivariate VAR(1) example process with cointegrating rank 1:

$$\begin{bmatrix} y_{1t} \\ y_{2t} \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} y_{1,t-1} \\ y_{2,t-1} \end{bmatrix} + \begin{bmatrix} u_{1t} \\ u_{2t} \end{bmatrix}. \tag{6.5.7}$$

The corresponding VECM representation is

$$\Delta y_t = -\begin{bmatrix} 1 & -1 \\ 0 & 0 \end{bmatrix} y_{t-1} + u_t = \begin{bmatrix} -1 \\ 0 \end{bmatrix} [1, -1] y_{t-1} + u_t,$$

that is,

$$\alpha = \begin{bmatrix} -1 \\ 0 \end{bmatrix}, \qquad \beta' = [1, -1].$$

For this process, it is easily seen that $\Phi_0 = I_2$ and

$$\Phi_j = A_1^j = \begin{bmatrix} 0 & 1 \\ 0 & 1 \end{bmatrix}, \qquad j = 1, 2, \ldots,$$

which implies

$$\Sigma_y(h) = \sum_{j=0}^{h-1} \Phi_j \Sigma_u \Phi_j' = \Sigma_u + (h-1) \begin{bmatrix} \sigma_2^2 & \sigma_2^2 \\ \sigma_2^2 & \sigma_2^2 \end{bmatrix}, \qquad h = 1, 2, \ldots,$$

where σ_2^2 is the variance of u_{2t}. Moreover, the conditional expectations are $y_{k,t}(h) = y_{2,t}$ $(k = 1, 2)$. Hence, the forecast intervals are

$$\left[y_{2,t} - z_{(\alpha/2)} \sqrt{\sigma_k^2 + (h-1)\sigma_2^2},\ y_{2,t} + z_{(\alpha/2)} \sqrt{\sigma_k^2 + (h-1)\sigma_2^2} \right], \quad k = 1, 2,$$

where $z_{(\alpha/2)}$ is the $(1-\frac{\alpha}{2})100$ percentage point of the standard normal distribution. It is easy to see that the length of this interval is unbounded for $h \to \infty$.

If there are cointegrated variables, some linear combinations can be forecasted with bounded forecast error variance, however. To see this, multiply (6.5.7) by

$$\begin{bmatrix} 1 & -1 \\ 0 & 1 \end{bmatrix}.$$

Thereby we get

$$\begin{bmatrix} 1 & -1 \\ 0 & 1 \end{bmatrix} y_t = \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} y_{t-1} + \begin{bmatrix} 1 & -1 \\ 0 & 1 \end{bmatrix} u_t,$$

which implies that the cointegration relation $z_t := y_{1t} - y_{2t} = u_{1t} - u_{2t}$ is zero mean white noise. Thus, the forecast intervals for z_t for any forecast horizon h are of constant length,

$$[z_t(h) - z_{(\alpha/2)}\sigma_z(h),\ z_t(h) + z_{(\alpha/2)}\sigma_z(h)] = [-z_{(\alpha/2)}\sigma_z,\ z_{(\alpha/2)}\sigma_z],$$

where $\sigma_z^2 := \mathrm{Var}(u_{1t}) + \mathrm{Var}(u_{2t}) - 2\mathrm{Cov}(u_{1t}, u_{2t})$ is the variance of z_t and $z_t(h) = 0$ for $h \geq 1$ has been used.

If deterministic terms are present, we may use the foregoing formulas for the mean-adjusted variables and then add the deterministic terms for the forecast period to the mean-adjusted forecasts. More precisely, if $y_t = \mu_t + x_t$, where μ_t is the deterministic term and x_t is the stochastic part, a forecast for y_{t+h} is obtained from a forecast $x_t(h)$ for x_{t+h} by simply adding μ_{t+h}, $y_t(h) = \mu_{t+h} + x_t(h)$. By the very nature of a deterministic term, μ_{t+h} is known, of course.

In practice, the parameters $A_1, \dots, A_p$, Σ_u, and and those of the deterministic part are usually unknown. The consequences of replacing them by estimators will be discussed in Chapter 7.

6.6 Causality Analysis

From the discussion in the previous subsection, it follows easily that the restrictions characterizing Granger-noncausality are exactly the same as in the stable case. More precisely, suppose that the vector y_t in (6.5.1) is partitioned in M- and $(K-M)$-dimensional subvectors z_t and x_t,

$$y_t = \begin{bmatrix} z_t \\ x_t \end{bmatrix} \quad \text{and} \quad A_i = \begin{bmatrix} A_{11,i} & A_{12,i} \\ A_{21,i} & A_{22,i} \end{bmatrix}, \quad i = 1, \dots, p,$$

where the A_i are partitioned in accordance with the partitioning of y_t. Then x_t does not Granger-cause z_t if and only if

$$A_{12,i} = 0, \quad i = 1, \dots, p. \tag{6.6.1}$$

In turn, z_t does not Granger-cause x_t if and only if $A_{21,i} = 0$ for $i = 1, \dots, p$. It is also easy to derive the corresponding restrictions for the VECM,

$$\begin{bmatrix} \Delta z_t \\ \Delta x_t \end{bmatrix} = \begin{bmatrix} \mathbf{\Pi}_{11} & \mathbf{\Pi}_{12} \\ \mathbf{\Pi}_{21} & \mathbf{\Pi}_{22} \end{bmatrix} \begin{bmatrix} z_{t-1} \\ x_{t-1} \end{bmatrix} + \sum_{i=1}^{p-1} \begin{bmatrix} \mathbf{\Gamma}_{11,i} & \mathbf{\Gamma}_{12,i} \\ \mathbf{\Gamma}_{21,i} & \mathbf{\Gamma}_{22,i} \end{bmatrix} \begin{bmatrix} \Delta z_{t-i} \\ \Delta x_{t-i} \end{bmatrix} + u_t,$$

where all matrices are partitioned in line with y_t. From (6.3.6) it follows immediately, that the restrictions in (6.6.1) can be written equivalently as

$$\mathbf{\Pi}_{12} = 0 \quad \text{and} \quad \mathbf{\Gamma}_{12,i} = 0 \quad \text{for } i = 1, \dots, p-1. \tag{6.6.2}$$

In other words, in order to check Granger-causality, we just have to test a set of linear hypotheses. It will be seen in the next chapter that in the case of cointegrated processes, testing these restrictions is not as straightforward as for stationary processes.

Also restrictions for multi-step causality and instantaneous causality can be placed on the VAR coefficients and the residual covariance matrix in the same way as in Chapter 2. Especially for the former restrictions, constructing valid asymptotic tests is not straightforward, however.

6.7 Impulse Response Analysis

Integrated and cointegrated systems must be interpreted cautiously. As mentioned in Section 6.3, in cointegrated systems the term $\beta' y_t$ is usually thought of as representing the long-run equilibrium relations between the variables. Suppose there is just one such relation, say

$$\beta_1 y_{1t} + \cdots + \beta_K y_{Kt} = 0,$$

or, if $\beta_1 \neq 0$,

$$y_{1t} = -\frac{\beta_2}{\beta_1} y_{2t} - \cdots - \frac{\beta_K}{\beta_1} y_{Kt}.$$

It is tempting to argue that the long-run effect of a unit increase in y_2 will be a change of size β_2/β_1 in y_1. This, however, ignores all the other relations between the variables which are summarized in a VAR(p) model or the corresponding VECM. A one-time unit innovation in y_2 may affect various other variables which also have an impact on y_1. Therefore, the long-run effect of a y_2-innovation on y_1 may be quite different from $-\beta_2/\beta_1$. The impulse responses may give a better picture of the relations between the variables.

In Chapter 2, Section 2.3.2, the impulse responses of stationary, stable VAR(p) processes were shown to be the coefficients of specific MA representations. An unstable, integrated or cointegrated VAR(p) process does not possess valid MA representations of the types discussed in Chapter 2. Yet the Φ_i and Θ_i matrices can be computed as in Section 2.3.2. For the Φ_i's we have seen this in Section 6.5 and, from the discussion in that section, it is easy to see that the elements of the $\Phi_i = (\phi_{jk,i})$ matrices may represent impulse responses just as in the stable case. More precisely, $\phi_{jk,i}$ represents the response of variable j to a unit forecast error in variable k, i periods ago, if the system reflects the actual responses to forecast errors. Recall that in stable processes the responses taper off to zero as $i \to \infty$. This property does not necessarily hold in unstable systems where the effect of a one-time impulse may not die out asymptotically.

In Section 2.3, we have also considered accumulated impulse responses, responses to orthogonalized residuals and forecast error variance decompositions. These tools for structural analysis are all available for unstable systems as well, using precisely the same formulas as in Chapter 2. The only quantities that cannot be computed in general are the total "long-run effects" or total multipliers Ψ_∞ and Ξ_∞ because they may not be finite.

To illustrate impulse response analysis of cointegrated systems, we consider the following VECM:

$$\begin{bmatrix} \Delta R_t \\ \Delta Dp_t \end{bmatrix} = \begin{bmatrix} -0.07 \\ 0.17 \end{bmatrix} (R_{t-1} - 4Dp_{t-1}) + \begin{bmatrix} 0.24 & -0.08 \\ 0 & -0.31 \end{bmatrix} \begin{bmatrix} \Delta R_{t-1} \\ \Delta Dp_{t-1} \end{bmatrix}$$

$$+ \begin{bmatrix} 0 & -0.13 \\ 0 & -0.37 \end{bmatrix} \begin{bmatrix} \Delta R_{t-2} \\ \Delta Dp_{t-2} \end{bmatrix} + \begin{bmatrix} 0.20 & -0.06 \\ 0 & -0.34 \end{bmatrix} \begin{bmatrix} \Delta R_{t-3} \\ \Delta Dp_{t-3} \end{bmatrix} + \begin{bmatrix} u_{1,t} \\ u_{2,t} \end{bmatrix}, \tag{6.7.1}$$

$$\Sigma_u = \begin{bmatrix} 2.61 & -0.15 \\ -0.15 & 2.31 \end{bmatrix} \times 10^{-5}$$

and the corresponding correlation matrix is

$$R_u = \begin{bmatrix} 1 & -0.06 \\ -0.06 & 1 \end{bmatrix}.$$

This model is from Lütkepohl (2004, Eq. (3.41)). The variables are a long-term interest rate (R_t) and the quarterly inflation rate (Dp_t). The coefficients are estimated from quarterly German data. Deterministic terms have been deleted because they are not important for the present analysis.

In contrast to the inflation/interest rate example system considered in Chapter 2, the two variables in the present system are $I(1)$. The cointegration relation, $R_t - 4Dp_t$, is just the real interest rate because $4Dp_t$ is the annual inflation rate and R_t is an annual nominal interest rate. Thus, in the present model the real interest rate is stationary. This relation is sometimes called the Fisher effect. The zero restrictions have been determined by a subset modelling algorithm. The residual covariance matrix is almost diagonal.

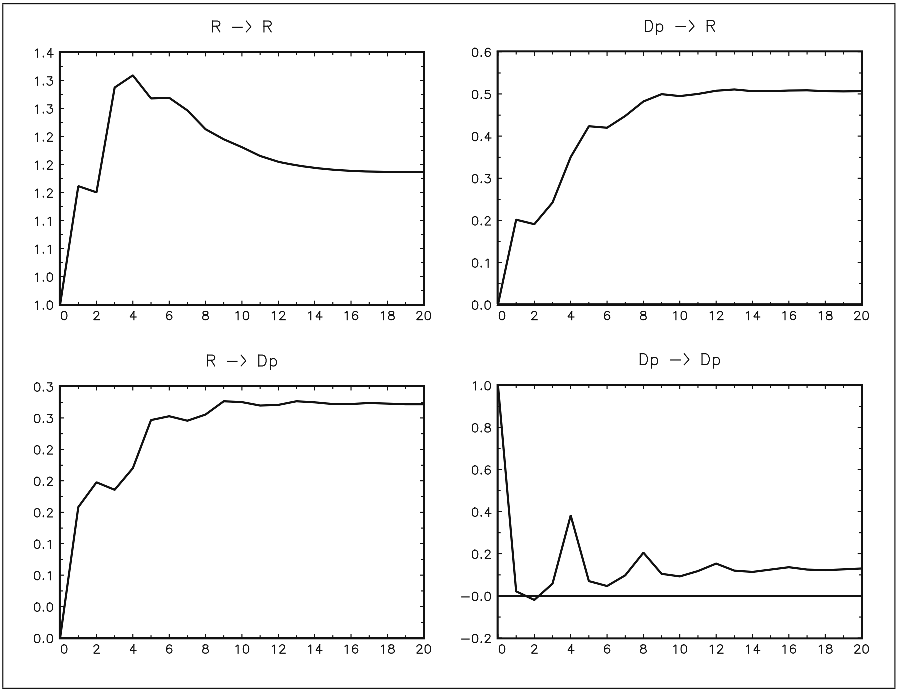

Fig. 6.4. Forecast error impulse responses of VECM (6.7.1).

Therefore, forecast error impulse responses should be similar to orthogonalized impulse responses, except for the scaling. The two types of impulse responses are shown in Figures 6.4 and 6.5, respectively. Indeed, the shape of corresponding impulse responses in the two figures is quite similar. A remarkable feature of the impulse responses is that they do not die out to zero when the time span after the impulse increases but approach some nonzero value. Clearly, this reflects the nonstationarity of the system where a one-time impulse can have permanent effects.

Using the orthogonalized impulse responses, it is also possible to compute forecast error variance decompositions based on the same formulas as in Chapter 2, Section 2.3.3. For the example system, they are shown in Figure 6.6. They look similar to forecast error variance decompositions from a stationary VAR process. Of course, there is no reason why they should look differently than in the stationary case.

As discussed in Chapter 2, interpreting the forecast error and orthogonalized impulse responses used here is often problematic if there is significant correlation between the components of the residuals u_t. It will be discussed in Chapter 9 how identifying restrictions for impulse responses can be imposed in the VECM framework.

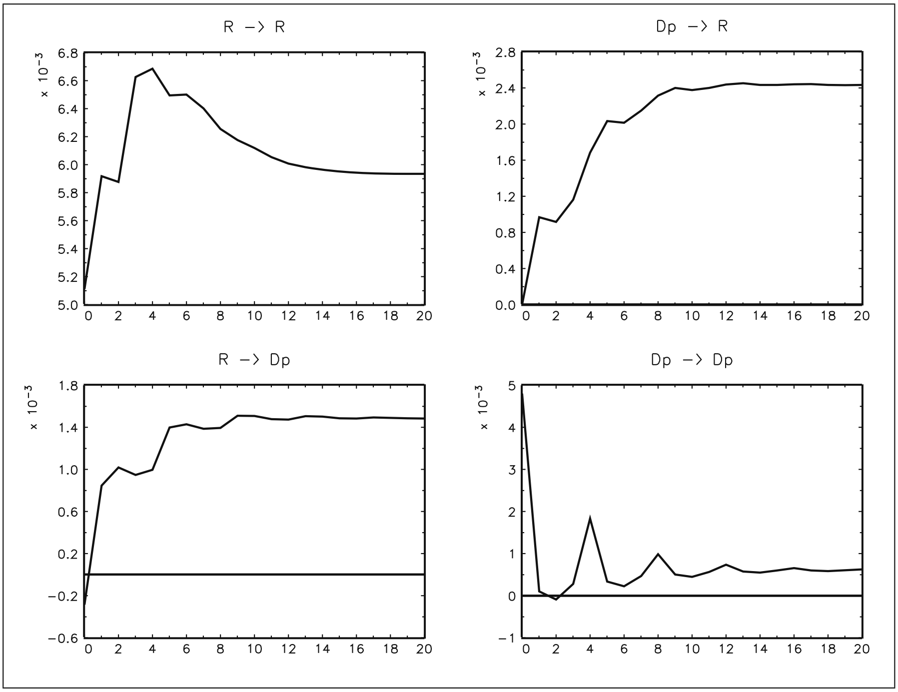

Fig. 6.5. Orthogonalized impulse responses of VECM (6.7.1).

6.8 Exercises

Problem 6.1
Consider the process

$$y_t = \begin{bmatrix} 1 & 0 \\ 0 & \psi \end{bmatrix} y_{t-1} + u_t$$

with residual covariance matrix

$$\Sigma_u = \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}.$$

(a) What is the cointegrating rank of the process?
(b) Write the process in VECM form.

Problem 6.2
Determine the roots of the reverse characteristic polynomial and, if applicable, the cointegrating rank of the process

$$y_t = \begin{bmatrix} 1.1 & -0.2 \\ -0.2 & 1.4 \end{bmatrix} y_{t-1} + u_t.$$

Can you write the process in VECM form?

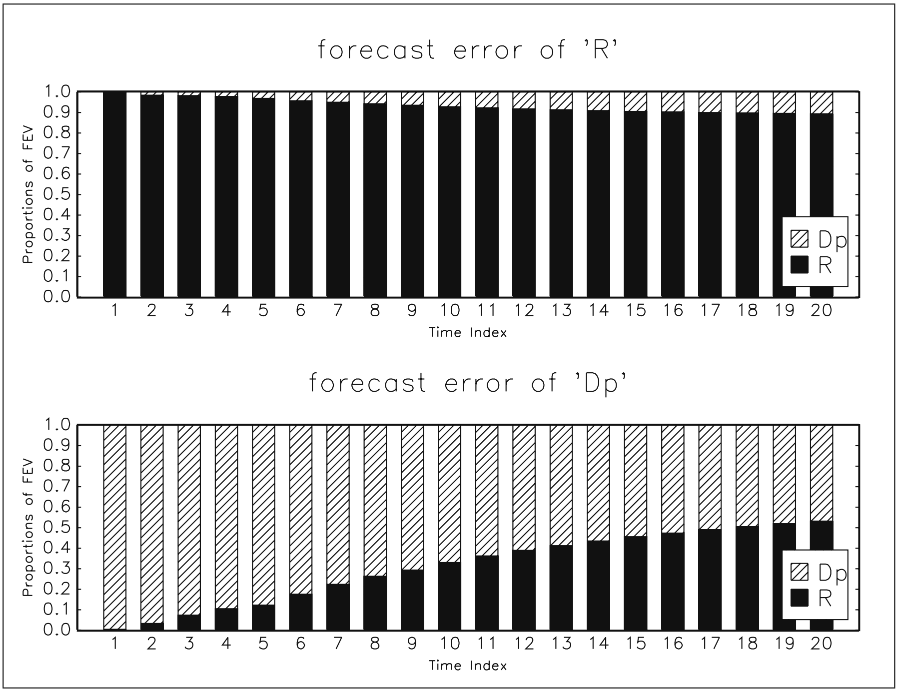

Fig. 6.6. Forecast error variance decomposition of VECM (6.7.1).

Problem 6.3
What is the maximum possible cointegrating rank of a three-dimensional process $y_t = (y_{1t}, y_{2t}, y_{3t})'$,

(a) if y_{1t}, y_{2t} are $I(0)$ and y_{3t} is $I(1)$?
(b) if y_{1t}, y_{2t}, and y_{3t} are $I(1)$ and y_{1t} and y_{2t} are not cointegrated in a bivariate system?
(c) if y_{1t}, y_{2t}, and y_{3t} are $I(1)$ and $(y_{1t}, y_{2t})'$ and $(y_{2t}, y_{3t})'$ are not cointegrated as bivariate systems?

Problem 6.4
Find the Beveridge-Nelson decomposition associated with the VECM

$$\Delta y_t = \alpha\beta' y_{t-1} + u_t,$$

(a) if all initial values are zero ($y_t = u_t = 0$ for $t \leq 0$),
(b) if y_0 is nonzero.

Problem 6.5
Derive the VECM form of y_t if the deterministic term is $\mu_t = \mu_0 + \delta I_{(t>T_B)}$, where $I_{(t>T_B)}$ is a shift dummy variable which is zero up to time T_B and then jumps to one and δ is the associated $(K \times 1)$ parameter vector.

Problem 6.6
Consider the quarterly process $y_t = \mu_t + x_t$, where x_t has a VECM representation as in (6.4.2) and

$$\mu_t = \mu_0 + \mu_1 t + \delta_1 s_{1t} + \delta_2 s_{2t} + \delta_3 s_{3t}.$$

Here μ_0, μ_1, δ_1, δ_2, and δ_3 are K-dimensional parameter vectors and the s_{it}'s $(i = 1, 2, 3)$ are seasonal dummy variables. Determine the VECM representation of y_t.

Problem 6.7
Consider the VECM

$$\Delta y_t = \begin{bmatrix} -0.1 \\ 0.1 \end{bmatrix} (1, -1) y_{t-1} + u_t.$$

(a) Rewrite the process in VAR form.
(b) Determine the roots of the reverse characteristic polynomial.
(c) Determine forecast intervals for the two variables for forecast horizon h.
(d) Has a forecast error impulse in y_{1t} a permanent impact on y_{2t}? Has a forecast error impulse in y_{2t} a permanent impact on y_{1t}?