3 Vector Error Correction Models

Many economic variables exhibit persistent upward or downward movement. This feature can be generated by stochastic trends in integrated variables. If the same stochastic trend is driving a set of integrated variables jointly, they are called cointegrated. In this case, certain linear combinations of integrated variables are stationary. Such linear combinations that link the variables to a common trend path are called cointegrating relationships. They sometimes may be interpreted as equilibrium relationships in economic models.

Cointegrating relationships can be imposed by reparameterizing the VAR model as a vector error correction model (VECM).[1] In Section 3.1 cointegrated variables are introduced and VECMs are set up. Sections 3.2 and 3.3 consider the estimation as well as the specification of VECMs. Diagnostic tools are presented in Section 3.4, and the implications of including cointegrated variables in VAR models for forecasting and Granger causality analysis are discussed in Section 3.5. Our focus in this chapter is on reduced-form models. We leave extensions to structural VECMs to later chapters.

The concept of cointegration was introduced in the econometrics literature by Granger (1981) and Engle and Granger (1987). Early work on error correction models goes back to Sargan (1964), Davidson, Hendry, Srba, and Yeo (1978), Hendry and von Ungern-Sternberg (1981), and Salmon (1982). Lütkepohl (1982b) discusses the cointegration feature without using the cointegration terminology. A full analysis of the VECM is presented in Johansen (1995), among others. Parts of the present chapter follow closely Lütkepohl (2005, part II; 2006, 2009).

3.1 Cointegrated Variables and Vector Error Correction Models

3.1.1 Common Trends and Cointegration

Cointegrated processes were introduced by Granger (1981) and Engle and Granger (1987). If two integrated variables share a common stochastic trend

[1] Some researchers also refer to these models as vector equilibrium correction models.

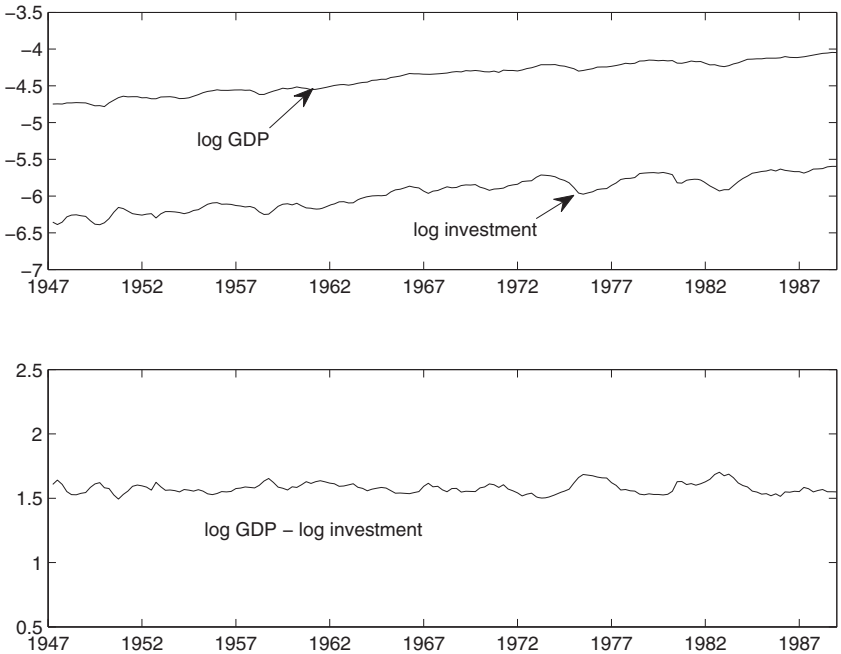

Figure 3.1. Logs of quarterly U.S. real GDP and real investment for 1947q1–1988q4.

such that a linear combination of these variables is stationary, they are called cointegrated. For example, the plots of quarterly U.S. log output and investment in the upper panel of Figure 3.1 both exhibit an upward trend. Because both series are driven by the same trend, the log of the GDP-investment ratio is fluctuating about a constant mean. As a result, the difference between the log series in the lower panel of Figure 3.1 has no obvious trend anymore. It is mean reverting and appears stationary.

The concept of cointegration may also be applied to linear combinations of more than two $I(1)$ variables. More formally, we say that a set of $I(1)$ time series variables is cointegrated if there exists a linear combination of these variables that is $I(0)$. Generalizing this concept to higher orders of integration, the variables in a K-dimensional process y_t are cointegrated if the components are $I(d)$ and there exists a linear combination $z_t = \beta' y_t$ with $\beta = (\beta_1, \ldots, \beta_K)' \neq 0$ such that z_t is $I(d^*)$ with $d^* < d$. The vector β is called a cointegrating vector or a cointegration vector. For example, let $z_t = p_t - p_t^* - e_t \sim I(0)$, where $p_t \sim I(1)$ denotes the log of the domestic consumer price index, $p_t^* \sim I(1)$ denotes the log of the foreign consumer price index, and $e_t \sim I(1)$ denotes the log of the nominal exchange rate expressed in domestic currency values per unit of foreign currency. Then $\beta = (1, -1, -1)'$ is a cointegrating vector. This

relationship embodies the view that under standard arbitrage conditions the real exchange rate must be $I(0)$, even when its components are not.

A cointegrating vector is not unique. In the real exchange rate example, multiplying β by any nonzero constant would result in another equally valid cointegrating vector. As a matter of convention, we typically normalize the coefficients of β such that one of the elements of β is 1. Note that the remaining values of the cointegrating vector in general do not have to be restricted to integers.

More generally, it is convenient to call a K-dimensional $I(d)$ process y_t cointegrated if there is a linear combination $\beta' y_t$ with $\beta \neq 0$ that is integrated of order less than d. Notice that this definition differs slightly from that given by Engle and Granger (1987) in that it also covers the case when the components of y_t have no common trend. For instance, if $y_t = (y_{1t}, y_{2t})'$ is a bivariate process such that $y_{1t} \sim I(1)$ and $y_{2t} \sim I(0)$, then the bivariate process y_t is $I(1)$ and has to be differenced once to make it stationary. However, the linear combination

$$(0, 1)\begin{pmatrix} y_{1t} \\ y_{2t} \end{pmatrix} = y_{2t}$$

is $I(0)$ and, hence, the process is cointegrated according to our slightly more general definition, although there is no common trend and, hence, there is no genuine cointegration in the original sense.

The VEC Model. In a system of variables, there may be several linearly independent cointegrating vectors. In that case linear combinations of these vectors are also cointegrating vectors because linear combinations of stationary variables are stationary. To embed the concept of cointegration in the VAR framework, suppose for the moment that all individual variables are $I(1)$ or $I(0)$ and the DGP is a K-dimensional VAR(p) process,

$$y_t = A_1 y_{t-1} + \cdots + A_p y_{t-p} + u_t, \tag{3.1.1}$$

without deterministic terms. Subtracting y_{t-1} on both sides of the equation and rearranging terms yields the VECM

$$\Delta y_t = \mathbf{\Pi} y_{t-1} + \mathbf{\Gamma}_1 \Delta y_{t-1} + \cdots + \mathbf{\Gamma}_{p-1} \Delta y_{t-p+1} + u_t, \tag{3.1.2}$$

where

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

and

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

Among the regressors in (3.1.2) the only nonstationary variable is y_{t-1}. Since the left-hand side of equation (3.1.2) is $I(0)$, so must be the right-hand side, which requires $\mathbf{\Pi} y_{t-1}$ to be $I(0)$.

Because the variables have unit roots individually,

$$\det(I_K - A_1 z - \cdots - A_p z^p) = 0$$

for $z = 1$, and, thus, the matrix $\mathbf{\Pi}$ is singular. Suppose this matrix has rank r. Then there are r linearly independent cointegrating relationships. Hence, the rank of $\mathbf{\Pi}$ is called the cointegration rank or cointegrating rank of the process y_t. Observe that any $K \times K$ matrix of rank r can be decomposed as a product of two $K \times r$ matrices of full column rank. Let α and β be two $K \times r$ matrices of rank r such that $\mathbf{\Pi} = \alpha\beta'$. Since (1) any linear transformation of stationary variables is stationary, since (2) $\mathbf{\Pi} y_t$ is stationary, and since (3) $(\alpha'\alpha)^{-1}\alpha'\mathbf{\Pi} y_t$ is a linear transformation of $\mathbf{\Pi} y_t$, the latter linear transformation, which equals $\beta' y_t$ after substituting $\mathbf{\Pi} = \alpha\beta'$, is also stationary, and, hence, the rows of β' are cointegration vectors. The matrix β' is therefore called the cointegrating matrix, and the matrix α is sometimes referred to as the loading matrix. Substituting the matrix $\alpha\beta'$ for $\mathbf{\Pi}$ in (3.1.2) yields

$$\Delta y_t = \alpha\beta' y_{t-1} + \mathbf{\Gamma}_1 \Delta y_{t-1} + \cdots + \mathbf{\Gamma}_{p-1} \Delta y_{t-p+1} + u_t, \tag{3.1.3}$$

which is called a VECM because it explicitly includes the lagged error correction (EC) term $\alpha\beta' y_{t-1}$.

Two special cases of this model merit further discussion. One special case arises when $r = K$, in which case the process is already stable in levels. All variables are $I(0)$ in levels, and there is no need to consider a VECM. The other special case arises when $r = 0$. In that case, the EC term is zero and Δy_t has a stable VAR$(p-1)$ representation in differences.

If the point of departure is the VECM (3.1.2), the corresponding levels VAR representation can be recovered easily by noting that

$$\begin{aligned} A_1 &= \mathbf{\Pi} + I_K + \mathbf{\Gamma}_1, \\ A_i &= \mathbf{\Gamma}_i - \mathbf{\Gamma}_{i-1}, \quad i = 2, \ldots, p-1, \\ A_p &= -\mathbf{\Gamma}_{p-1}. \end{aligned} \tag{3.1.4}$$

As mentioned earlier, cointegration relationships are not unique. This fact is reflected in the nonuniqueness of the decomposition of the $K \times K$ matrix $\mathbf{\Pi} = \alpha\beta'$. Any nonsingular $r \times r$ matrix Q gives rise to a decomposition $\mathbf{\Pi} = \alpha^*\beta^{*\prime}$, where $\alpha^* = \alpha Q'$ and $\beta^* = \beta Q^{-1}$. A convenient normalisation relies on the fact that β can always be chosen as

$$\beta = \begin{bmatrix} I_r \\ \beta_{(K-r)} \end{bmatrix}, \tag{3.1.5}$$

where $\beta_{(K-r)}$ is $(K-r)\times r$, possibly after the variables have been rearranged suitably. For example, if the cointegrating rank is 1, all cointegrating relationships are multiples of a single relationship, and the normalization in (3.1.5) implies that this relationship can be written as a single equation with one variable on the left-hand side and the others on the right-hand side. Using $\beta = (1, \beta'_{(K-1)})'$ with $\beta'_{(K-1)} = (\beta_{(K-1),1}, \ldots, \beta_{(K-1),K-1})'$, yields

$$y_{1t} = -\beta_{(K-r),1} y_{2t} - \cdots - \beta_{(K-r),K-1} y_{Kt} + z_t,$$

where $z_t \sim I(0)$. Usually we think of cointegration relationships as linear combinations of $I(1)$ variables. If some of the elements of y_t are $I(0)$ in levels, there is an additional cointegrating relationship for each stationary component of y_t. Because these cointegrating vectors are linearly independent columns of β, the cointegrating rank must be at least as large as the number of $I(0)$ variables in the system.

The Granger Representation of the VECM. Granger's representation theorem, as stated by Johansen (1995, Theorem 4.2), is another useful representation of cointegrated processes. It may be stated using the orthogonal complements of matrices. For $m \geq n$, an orthogonal complement of the $m \times n$ matrix M with $\text{rank}(M) = n$ is denoted by $M_\perp$. Put differently, $M_\perp$ is any $m \times (m-n)$ matrix with $\text{rank}(M_\perp) = m - n$ and $M'M_\perp = 0$. Note that the $m \times m$ matrix $[M, M_\perp]$ is nonsingular. If M is a nonsingular square matrix ($m = n$), then $M_\perp = 0$, and, if $M = 0$, then $M_\perp = I_m$.

Let y_t be a K-dimensional cointegrated $I(1)$ process as in (3.1.3) with cointegrating rank r, $0 \leq r < K$. Then it can be shown that

$$y_t = \Xi \sum_{i=1}^{t} u_i + \Xi^*(L) u_t + y_0^*, \tag{3.1.6}$$

where

$$\Xi = \beta_\perp \left[\alpha'_\perp \left(I_K - \sum_{i=1}^{p-1} \Gamma_i \right) \beta_\perp \right]^{-1} \alpha'_\perp, \tag{3.1.7}$$

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

The representation in (3.1.6) is a multivariate version of the Beveridge-Nelson decomposition of y_t discussed in Chapter 2 and is known as the Granger representation of the process. It is also known as the common-trends representation. Equation (3.1.6) decomposes the process y_t into $I(1)$ and $I(0)$ components. The term $\sum_{i=1}^{t} u_i$ is a K-dimensional random walk. However, the matrix Ξ has rank $K - r$. For Ξ to be properly defined in (3.1.7), 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. Hence, $\operatorname{rank}(\mathbf{\Xi})=K-r$ and, thus, the term $\mathbf{\Xi}\sum_{i=1}^{t}u_i$ on the right-hand side of (3.1.6) effectively consists of a $(K-r)$-dimensional random walk or common trends component. Consequently, y_t is driven by $K-r$ common trends.

3.1.2 Deterministic Terms in Cointegrated Processes

Deterministic terms complicate the specification of integrated processes. For example, a constant term in a random walk process generates a linear trend in the mean, as seen in Chapter 2. This linear trend is distinct from the stochastic trend implied by the random walk component. Therefore, it is necessary to pay special attention to the implications of deterministic terms in cointegrated processes. Suppose that

$$y_t=\mu_t+x_t, \tag{3.1.8}$$

where x_t is a K-dimensional zero mean VAR(p) process with possibly cointegrated variables and μ_t is a $K\times 1$ deterministic term. In practice, μ_t often includes a constant and possibly an additional deterministic time trend. In other words, $\mu_t=\mu_0$ or $\mu_t=\mu_0+\mu_1 t$, where μ_0 and μ_1 are fixed K-dimensional parameter vectors.

The VECM with Intercept. The DGP of x_t is assumed to be

$$\begin{aligned}\Delta x_t&=\alpha\beta' x_{t-1}+\mathbf{\Gamma}_1\Delta x_{t-1}+\cdots+\mathbf{\Gamma}_{p-1}\Delta x_{t-p+1}+u_t\\&=\mathbf{\Pi}x_{t-1}+\mathbf{\Gamma}_1\Delta x_{t-1}+\cdots+\mathbf{\Gamma}_{p-1}\Delta x_{t-p+1}+u_t.\end{aligned} \tag{3.1.9}$$

For $\mu_t=\mu_0$, we have $x_t=y_t-\mu_0$ such that $\Delta y_t=\Delta x_t$ and, hence,

$$\begin{aligned}\Delta y_t&=\alpha\beta'(y_{t-1}-\mu_0)+\mathbf{\Gamma}_1\Delta y_{t-1}+\cdots+\mathbf{\Gamma}_{p-1}\Delta y_{t-p+1}+u_t\\&=\alpha\beta^{o\prime}\begin{pmatrix}y_{t-1}\\1\end{pmatrix}+\mathbf{\Gamma}_1\Delta y_{t-1}+\cdots+\mathbf{\Gamma}_{p-1}\Delta y_{t-p+1}+u_t\\&=\mathbf{\Pi}^o y_{t-1}^o+\mathbf{\Gamma}_1\Delta y_{t-1}+\cdots+\mathbf{\Gamma}_{p-1}\Delta y_{t-p+1}+u_t,\end{aligned} \tag{3.1.10}$$

where $\beta^{o\prime}=[\beta',\delta']$ with $\delta'=-\beta'\mu_0$ being an $r\times 1$ vector,

$$y_{t-1}^o=\begin{pmatrix}y_{t-1}\\1\end{pmatrix}$$

and $\mathbf{\Pi}^o=[\mathbf{\Pi},\vartheta]$ is a $K\times(K+1)$ matrix with $\vartheta=-\mathbf{\Pi}\mu_0=\alpha\delta'$. Thus, a constant mean term in the additive representation (3.1.8) becomes an intercept

term in the cointegration relationship. In the implied VECM for Δy_t,

$$\begin{aligned}\Delta y_t &= \nu_0 + \alpha\beta' y_{t-1} + \boldsymbol{\Gamma}_1 \Delta y_{t-1} + \cdots + \boldsymbol{\Gamma}_{p-1}\Delta y_{t-p+1} + u_t \\ &= \nu_0 + \boldsymbol{\Pi} y_{t-1} + \boldsymbol{\Gamma}_1 \Delta y_{t-1} + \cdots + \boldsymbol{\Gamma}_{p-1}\Delta y_{t-p+1} + u_t, \quad (3.1.11)\end{aligned}$$

the intercept ν_0 has to satisfy the restrictions $\nu_0 = \alpha\delta'$. As a result, the intercept can be absorbed into the cointegration relationship. This fact ensures that the intercept does not generate a linear time trend in the mean of the y_t variables, consistent with our assumption that none of the model variables exhibits a linear time trend.

The VECM with Intercept and Trend. If $\mu_t = \mu_0 + \mu_1 t$ is a linear trend, we have $x_t = y_t - \mu_0 - \mu_1 t$ and $\Delta x_t = \Delta y_t - \mu_1$. Thus, substituting in (3.1.9) yields

$$\begin{aligned}\Delta y_t - \mu_1 = \alpha\beta'(y_{t-1} - \mu_0 - \mu_1(t-1)) + \boldsymbol{\Gamma}_1(\Delta y_{t-1} - \mu_1) + \cdots \\ + \boldsymbol{\Gamma}_{p-1}(\Delta y_{t-p+1} - \mu_1) + u_t. \quad (3.1.12)\end{aligned}$$

Rearranging the deterministic terms we have

$$\begin{aligned}\Delta y_t &= \nu + \alpha[\beta', \eta'] \begin{pmatrix} y_{t-1} \\ t-1 \end{pmatrix} + \boldsymbol{\Gamma}_1 \Delta y_{t-1} + \cdots + \boldsymbol{\Gamma}_{p-1}\Delta y_{t-p+1} + u_t \\ &= \nu + \boldsymbol{\Pi}^+ y_{t-1}^+ + \boldsymbol{\Gamma}_1 \Delta y_{t-1} + \cdots + \boldsymbol{\Gamma}_{p-1}\Delta y_{t-p+1} + u_t, \quad (3.1.13)\end{aligned}$$

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

$$y_{t-1}^+ = \begin{pmatrix} y_{t-1} \\ t-1 \end{pmatrix}.$$

In this representation the intercept term ν is left unrestricted, whereas the linear trend term can be absorbed into the cointegration relationships.

Finally, it can be shown that in the model

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

with unrestricted linear trend term, $I(1)$ variables actually generate quadratic deterministic trends in y_t.

It is also possible that the trend slope parameter μ_1 is orthogonal to the cointegration matrix such that $\beta'\mu_1 = 0$. In that case, $\eta = 0$, and there is no linear trend term in the cointegrating relationships, although the individual variables have linear trends. The model

$$\begin{aligned}\Delta y_t &= \nu + \alpha\beta' y_{t-1} + \boldsymbol{\Gamma}_1 \Delta y_{t-1} + \cdots + \boldsymbol{\Gamma}_{p-1}\Delta y_{t-p+1} + u_t \\ &= \nu + \boldsymbol{\Pi} y_{t-1} + \boldsymbol{\Gamma}_1 \Delta y_{t-1} + \cdots + \boldsymbol{\Gamma}_{p-1}\Delta y_{t-p+1} + u_t, \quad (3.1.14)\end{aligned}$$

with unrestricted intercept term ν is popular in applied work because a trend in the cointegrating relationships is sometimes regarded as implausible. Note that cointegrating relationships occasionally may be interpreted as equilibrium relationships. In that case, it is particularly implausible that the variables are driven apart by a deterministic trend. It is worth emphasizing, however, that, when there is no trend in the cointegrating relationships, the cointegration rank must be smaller than K in order to obtain a linear trend in the variables. If $r = K$, the process is stationary and, hence, a constant term in the model does not generate a linear trend.

Finally it is worth noting that the additive form of the drifting process in (3.1.8) facilitates the derivation of a common trends representation of y_t processes with deterministic terms. This representation may be obtained from the Granger representation (3.1.6) of x_t by adding the deterministic term, resulting in

$$y_t = \Xi \sum_{i=1}^{t} u_i + \Xi^*(L)u_t + y_0^* + \mu_t, \tag{3.1.15}$$

where all symbols are defined as in expression (3.1.6) and μ_t is the deterministic term defined in (3.1.8).

3.2 Estimation of VARs with Integrated Variables

VECMs are a convenient reparameterization of VAR models in levels. Clearly, from the estimator of the VECM parameters one can derive an estimator of the parameters of the corresponding VAR in levels. Alternatively, one can estimate the VAR model in levels directly, as already mentioned in the previous chapter. In this subsection, we discuss alternative estimation approaches for VECMs and compare their properties.

3.2.1 The VAR(1) Case

It is instructive to start with a simple VAR(1) model without any deterministic terms. Detailed derivations for this case are provided in Lütkepohl (2005, section 7.1). Here we just summarize and discuss the results. Consider the model

$$y_t = A_1 y_{t-1} + u_t, \tag{3.2.1}$$

where u_t is iid white noise with nonsingular covariance matrix Σ_u, i.e., $u_t \overset{iid}{\sim} (0, \Sigma_u)$. The corresponding VECM is

$$\Delta y_t = \Pi y_{t-1} + u_t = \alpha\beta' y_{t-1} + u_t, \tag{3.2.2}$$

where $\Pi = A_1 - I_K$ and $\text{rank}(\Pi) = \text{rank}(\alpha) = \text{rank}(\beta) = r$.

LS Estimation. If the cointegrating rank r is unknown and the A_1 matrix, or equivalently $\mathbf{\Pi}$, is estimated by LS, we have

$$\widehat{A}_1 - A_1 = \widehat{\mathbf{\Pi}} - \mathbf{\Pi} = \left(\sum_{t=1}^{T} u_t y_{t-1}'\right)\left(\sum_{t=1}^{T} y_{t-1} y_{t-1}'\right)^{-1}.$$

The asymptotic distribution of the estimators depends on the cointegrating rank. For $r = 0$, the process consists of K random walks, and the asymptotic distribution is

$$T(\widehat{A}_1 - A_1) \xrightarrow{d} \Sigma_u^{1/2}\left\{\int_0^1 \mathbf{W}_K d\mathbf{W}_K'\right\}'\left\{\int_0^1 \mathbf{W}_K \mathbf{W}_K' ds\right\}^{-1}\Sigma_u^{-1/2},$$

where $\mathbf{W}_K$ denotes a K-dimensional standard Brownian motion and $\Sigma_u^{1/2}$ is the square-root matrix of Σ_u.[2] Note that we have written the asymptotic distribution in matrix form, which is different from our previous notation which expressed multivariate distributions in vector form. This change of notation is of no consequence. The important point is that in this model a well-defined asymptotic distribution is obtained upon standardizing the estimator by T rather than $\sqrt{T}$. In other words, the estimator converges at a faster rate than in the stationary case. Moreover, its asymptotic distribution is not Gaussian. For univariate AR processes this situation is well known from the literature on Dickey-Fuller tests for unit roots. Our analysis in this subsection generalizes this result to the multivariate case.

If the cointegrating rank $r > 0$ and the cointegrating matrix β is normalized such that

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

only the elements of the $(K - r) \times r$ matrix $\beta_{(K-r)}$ are unknown. These elements can be estimated consistently. Given the normalization,

$$\mathbf{\Pi} = [\alpha, \alpha\beta_{(K-r)}'],$$

where the matrix $\mathbf{\Pi}$ is constructed by concatenating the matrices α and $\alpha\beta_{(K-r)}'$, one can express model (3.2.2) as

$$\Delta y_t - \alpha y_{t-1}^{(1)} = \alpha\beta_{(K-r)}' y_{t-1}^{(2)} + u_t = (y_{t-1}^{(2)\prime} \otimes \alpha)\mathrm{vec}(\beta_{(K-r)}') + u_t, \tag{3.2.3}$$

where $y_{t-1}^{(1)}$ and $y_{t-1}^{(2)}$ consist of the first r and the last $K - r$ elements of y_{t-1}, respectively.

[2] A Brownian motion is a continuous-time stochastic process with independent Gaussian increments that is commonly used to characterize limiting distributions of statistics that depend on integrated processes. For a formal definition of Brownian motions see Hamilton (1994).

GLS Estimation. Model (3.2.3) is a multivariate regression model with potentially different regressors in different equations. Therefore, GLS estimation may be more efficient than LS estimation. The GLS estimator of $\beta'_{(K-r)}$ can be shown to be

$$\widehat{\beta}'_{(K-r)} = (\alpha'\Sigma_u^{-1}\alpha)^{-1}\alpha'\Sigma_u^{-1} \times \left(\sum_{t=1}^{T}(\Delta y_t - \alpha y_{t-1}^{(1)})y_{t-1}^{(2)\prime}\right)\left(\sum_{t=1}^{T} y_{t-1}^{(2)}y_{t-1}^{(2)\prime}\right)^{-1}. \tag{3.2.4}$$

Note that this GLS estimator does not reduce to the LS estimator because the regressors differ across equations. Since α and Σ_u are unknown in practice, these quantities have to be replaced by estimates. Consistent estimators of α and Σ_u can be obtained from LS estimation of model (3.2.2). Given our normalization of β, the first r columns of the LS estimator $\widehat{\mathbf{\Pi}}$ are a consistent estimator of α, and Σ_u can be estimated in the usual way from the LS residuals.

Using these estimators in the expression for $\widehat{\beta}'_{(K-r)}$ in (3.2.4), we obtain a feasible GLS estimator that converges at rate T. In fact,

$$T(\widehat{\beta}'_{(K-r)} - \beta'_{(K-r)})$$

has a mixed normal asymptotic distribution that can be used for constructing valid asymptotic tests for the coefficients.

Using these estimators of α and β, we can estimate $\mathbf{\Pi}$ as $\widehat{\mathbf{\Pi}} = [\widehat{\alpha}, \widehat{\alpha}\widehat{\beta}'_{(K-r)}]$, which allows the construction of the estimator $\widehat{A}_1 = \widehat{\mathbf{\Pi}} + I_K$ of A_1. The convergence rate of this estimator is $\sqrt{T}$, and its asymptotic distribution is Gaussian. It can be shown that

$$\sqrt{T}\text{vec}(\widehat{A}_1 - A_1) = \sqrt{T}\text{vec}(\widehat{\mathbf{\Pi}} - \mathbf{\Pi}) \xrightarrow{d} \mathcal{N}(0, \beta\Gamma^{-1}\beta' \otimes \Sigma_u), \tag{3.2.5}$$

where $\Gamma \equiv \text{plim } T^{-1}\sum_{t=1}^{T}\beta' y_{t-1}y'_{t-1}\beta$ is a nonsingular $r \times r$ matrix.

LS Estimation with Known Cointegrating Matrix. In fact, the same asymptotic distribution is obtained when the cointegrating matrix β is known and only α is estimated by LS from the model

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

Denoting the resulting estimator by $\widehat{\alpha}$, where

$$\widehat{\alpha} = \left(\sum_{t=1}^{T}\Delta y_t y'_{t-1}\beta\right)\left(\sum_{t=1}^{T}\beta' y_{t-1}y'_{t-1}\beta\right)^{-1},$$

the corresponding estimator $\widehat{A}_1 = \widehat{\alpha}\beta' + I_K$ has precisely the asymptotic distribution given in expression (3.2.5). Thus, knowing the cointegrating rank r or the cointegrating matrix β does not improve the asymptotic efficiency of the LS estimator of A_1 or $\mathbf{\Pi}$. This result is a direct consequence of the faster convergence rate of the estimator of the cointegrating parameters. In fact, the same asymptotic distribution is obtained if the cointegrating rank is unknown and A_1 or $\mathbf{\Pi}$ are estimated by LS without accounting for the cointegrating rank. Thus, in a model with $r > 0$, knowing the true cointegrating rank is no advantage in estimation, as long as only asymptotic results are of interest. This purely asymptotic argument, however, ignores that not using all available information about the cointegrating structure of the model variables reduces the accuracy of the estimator in small samples.

The possible presence of cointegrated variables in the VAR model also may complicate statistical inference. An important feature of the asymptotic distribution in expression (3.2.5) is the singularity of the covariance matrix if $r < K$. This feature implies that standard inference will not be valid in general. For example, confidence intervals and t-tests for the coefficients may be misleading when they are based on the usual normal asymptotics. Clearly, the matrix $\beta\Gamma^{-1}\beta'$ depends on the cointegrating structure of the variables because it depends on the cointegrating matrix β. Suppose, for example, that the cointegrating rank is $r = 1$ and that y_t consists of the two components y_{1t} and y_{2t} such that $\beta' = (\beta_1, \beta_2)$. The null hypothesis for testing Granger causality from y_{1t} to y_{2t}, for example, is

$$\mathbb{H}_0 : a_{21,1} = 0.$$

This hypothesis can be tested with the t-ratio of $a_{21,1}$ obtained from (3.2.5) if the corresponding variance in the asymptotic covariance matrix is nonzero. The latter condition, however, requires that $\beta_1 \neq 0$. If both components of y_t are $I(1)$, there is no problem, and the t-ratio can be safely used for testing $\mathbb{H}_0$ because the cointegration relationship must necessarily involve both variables, and, thus, β_1 and β_2 are both nonzero. This fact was emphasized in Lütkepohl and Reimers (1992b). If, however, y_{2t} happens to be $I(0)$, then $\beta_1 = 0$, and a t-test based on standard normal critical values is not valid. This example shows that knowing the true cointegration structure (or at least some aspects of this structure such as the order of integration of both variables) can be important for valid inference in cointegrated VAR models. If the cointegration structure is not known, one option is to determine the cointegration properties of the data by statistical procedures. The other option is to estimate the model in levels, as discussed in Section 3.2.3.

FIML Estimation of Gaussian Processes with Known Cointegrating Rank. For a given rank r, it is also possible to estimate the VECM (3.2.2) without

the standardization of the cointegration matrix β in (3.1.5). In that case, one estimates β by the reduced-rank (RR) regression or canonical correlation procedure of Johansen (1988). Johansen proposes minimizing the determinant

$$\det\left(T^{-1}\sum_{t=1}^{T}(\Delta y_t - \mathbf{\Pi} y_{t-1})(\Delta y_t - \mathbf{\Pi} y_{t-1})'\right)$$

subject to the rank restriction $\text{rank}(\mathbf{\Pi}) = r$ or, equivalently, to minimize the determinant

$$\det\left(T^{-1}\sum_{t=1}^{T}(\Delta y_t - \alpha\beta' y_{t-1})(\Delta y_t - \alpha\beta' y_{t-1})'\right)$$

with respect to the $K \times r$ matrices α and β. The solution to this problem is based on the ordered eigenvalues $\lambda_1 \geq \cdots \geq \lambda_K$ and associated orthonormal eigenvectors $\eta_1, \ldots, \eta_K$ of the matrix

$$\left(\sum_{t=1}^{T} y_{t-1}y_{t-1}'\right)^{-1/2}\left(\sum_{t=1}^{T} y_{t-1}\Delta y_t'\right)\left(\sum_{t=1}^{T}\Delta y_t \Delta y_t'\right)^{-1}$$

$$\times\left(\sum_{t=1}^{T}\Delta y_t y_{t-1}'\right)\left(\sum_{t=1}^{T} y_{t-1}y_{t-1}'\right)^{-1/2}.$$

The resulting estimators are

$$\widetilde{\beta}' = [\eta_1, \ldots, \eta_r]'\left(\sum_{t=1}^{T} y_{t-1}y_{t-1}'\right)^{-1/2}$$

and

$$\widetilde{\alpha} = \left(\sum_{t=1}^{T}\Delta y_t y_{t-1}'\widetilde{\beta}\right)\left(\sum_{t=1}^{T}\widetilde{\beta}' y_{t-1}y_{t-1}'\widetilde{\beta}\right)^{-1}.$$

The procedure is equivalent to maximizing the log-likelihood function for a model with Gaussian residuals, $u_t \overset{iid}{\sim} \mathcal{N}(0, \Sigma_u)$. The estimators $\widetilde{\alpha}$ and $\widetilde{\beta}$ are not consistent because they are not separately identified. However, the corresponding estimator $\widetilde{\mathbf{\Pi}} = \widetilde{\alpha}\widetilde{\beta}'$ for $\mathbf{\Pi}$ is consistent and has the same asymptotic distribution as the LS estimator in expression (3.2.5).

3.2.2 Estimation of VECMs

Now suppose that the cointegrating rank r is greater than zero and consider a more general VECM without deterministic terms, but of lag order $p > 1$ such

that

$$\begin{aligned}\Delta y_t &= \alpha\beta' y_{t-1} + \mathbf{\Gamma}_1 \Delta y_{t-1} + \cdots + \mathbf{\Gamma}_{p-1}\Delta y_{t-p+1} + u_t \\ &= \alpha\beta' y_{t-1} + \mathbf{\Gamma} X_{t-1} + u_t, \end{aligned} \tag{3.2.6}$$

where

$$\mathbf{\Gamma} = [\mathbf{\Gamma}_1, \ldots, \mathbf{\Gamma}_{p-1}], \quad \text{and} \quad X_{t-1} = \begin{pmatrix} \Delta y_{t-1} \\ \vdots \\ \Delta y_{t-p+1} \end{pmatrix}.$$

The estimation of this model involves three steps. It is useful to begin by concentrating out $\mathbf{\Gamma}$, which allows us to focus on the problem of estimating $\alpha\beta'$. Given $\alpha\beta'$, the LS estimator of $\mathbf{\Gamma}$ is known to be

$$\widehat{\mathbf{\Gamma}}(\alpha\beta') = \left(\sum_{t=1}^{T}(\Delta y_t - \alpha\beta' y_{t-1})X_{t-1}'\right)\left(\sum_{t=1}^{T} X_{t-1}X_{t-1}'\right)^{-1}. \tag{3.2.7}$$

Replacing $\mathbf{\Gamma}$ in (3.2.6) by this estimator yields the concentrated model

$$R_{0t} = \alpha\beta' R_{1t} + u_t^*, \tag{3.2.8}$$

where

$$R_{0t} = \Delta y_t - \left(\sum_{t=1}^{T}\Delta y_t X_{t-1}'\right)\left(\sum_{t=1}^{T} X_{t-1}X_{t-1}'\right)^{-1} X_{t-1}$$

and

$$R_{1t} = y_{t-1} - \left(\sum_{t=1}^{T} y_{t-1}X_{t-1}'\right)\left(\sum_{t=1}^{T} X_{t-1}X_{t-1}'\right)^{-1} X_{t-1}.$$

It is not difficult to see that R_{0t} and R_{1t} are just the residuals from regressing Δy_t and y_{t-1}, respectively, on X_{t-1}. In the second step, ML and GLS methods are used to estimate α and β from the concentrated model (3.2.8). In the last step, the estimator of $\mathbf{\Gamma}$ is obtained by substituting the estimators of α and β into expression (3.2.7).

ML Estimation for Gaussian Processes. The parameters α and β can be estimated by RR regression as proposed by Johansen (1988) (see also Anderson 1951). As in the VAR(1) case considered earlier, this estimation method is

equivalent to ML estimation if the process is Gaussian. The concentrated log-likelihood function is

$$\log l = -\frac{KT}{2}\log(2\pi) - \frac{T}{2}\log(\det(\Sigma_u)) - \frac{1}{2}\mathrm{tr}\left[\sum_{t=1}^{T}(R_{0t} - \alpha\beta' R_{1t})'\Sigma_u^{-1}(R_{0t} - \alpha\beta' R_{1t})\right] \tag{3.2.9}$$

and the RR or ML estimators are

$$\widetilde{\beta}' = [\eta_1, \ldots, \eta_r]' S_{11}^{-1/2} \quad \text{and} \quad \widetilde{\alpha} = S_{01}\widetilde{\beta}(\widetilde{\beta}' S_{11}\widetilde{\beta})^{-1}, \tag{3.2.10}$$

where tr denotes the trace of a matrix, $S_{ij} = \sum_{t=1}^{T} R_{it}R_{jt}'/T$, $i = 0, 1$, and $\eta_1, \ldots, \eta_K$ are the orthonormal eigenvectors of the matrix $S_{11}^{-1/2}S_{10}S_{00}^{-1}S_{01}S_{11}^{-1/2}$ corresponding to its eigenvalues in nonincreasing order, $\lambda_1 \geq \cdots \geq \lambda_K$.

Implementation of the Johansen procedure does not require consistent estimation of the cointegrating matrix and hence does not require any normalization of the cointegrating matrix. Such normalizations are not required for many applications of VECMs such as impulse response analysis. They are only required if we are specifically interested in the cointegrating vectors. If β is normalized such that $\beta' = [I_r, \beta_{(K-r)}']$ as in (3.1.5) and the RR/ML estimator is normalized accordingly, then, under general conditions, $T\mathrm{vec}(\widetilde{\beta}_{(K-r)}' - \beta_{(K-r)}')$ converges in distribution to a Gaussian mixture distribution (Johansen 1995 or Lütkepohl 2005, chapter 7). Asymptotic mixed normality of the cointegration parameter estimator means that inference can be conducted similarly to asymptotically normal estimators.

The ML estimator of α given in (3.2.10) is the LS estimator obtained from the multivariate regression model

$$R_{0t} = \alpha\widetilde{\beta}' R_{1t} + u_t^*$$

with regressor matrix $\widetilde{\beta}' R_{1t}$. If $\widetilde{\beta}$ is replaced with a superconsistent estimator, the asymptotic properties of the corresponding $\widetilde{\alpha}$ estimator are standard and so are those of

$$\widetilde{\mathbf{\Gamma}}(\widetilde{\alpha}\widetilde{\beta}') = \left(\sum_{t=1}^{T}(\Delta y_t - \widetilde{\alpha}\widetilde{\beta}' y_{t-1})X_{t-1}'\right)\left(\sum_{t=1}^{T} X_{t-1}X_{t-1}'\right)^{-1}. \tag{3.2.11}$$

Under general conditions, these estimators converge at the usual $\sqrt{T}$ rate to an asymptotic normal distribution. It can be shown that

$$\sqrt{T}\mathrm{vec}([\widetilde{\alpha}, \widetilde{\mathbf{\Gamma}}] - [\alpha, \mathbf{\Gamma}]) \xrightarrow{d} \mathcal{N}(0, \Sigma_{\alpha,\mathbf{\Gamma}}),$$

where

$$\Sigma_{\alpha,\mathbf{r}} = \text{plim}\, T \begin{bmatrix} \beta' \sum_{t=1}^{T} y_{t-1} y_{t-1}' \beta & \beta' \sum_{t=1}^{T} y_{t-1} X_{t-1}' \\ \sum_{t=1}^{T} X_{t-1} y_{t-1}' \beta & \sum_{t=1}^{T} X_{t-1} X_{t-1}' \end{bmatrix}^{-1} \otimes \Sigma_u.$$

We illustrate the Johansen approach to estimating VECMs based on the same trivariate example already used in the context of the estimation of stationary VAR models. Recall that $y_t = (\Delta gnp_t, i_t, \Delta p_t)'$. We treat i_t and Δp_t as individually $I(1)$, but cointegrated. For expository purposes we also assume that the log of U.S. real GNP is not cointegrated with any of the other model variables. This implies that the cointegrating rank is 2 with $\Delta gnp_t \sim I(0)$ and, hence, trivially cointegrated with itself. The model includes an intercept as in (3.1.14). As before, we impose a lag order of $p = 4$, implying the existence of three augmented lags in the VECM representation. The ML estimates are

$$\widetilde{\nu} = \begin{bmatrix} 0.7520 \\ -0.3416 \\ -0.0564 \end{bmatrix},$$

$$\widetilde{\alpha} = \begin{bmatrix} -0.2617 & 0.1086 \\ 0.2470 & 0.1187 \\ 0.0353 & 0.0153 \end{bmatrix}, \quad \widetilde{\beta}' = \begin{bmatrix} 2.1932 & -0.0594 & 0.8530 \\ -0.3637 & -0.4489 & 1.4588 \end{bmatrix},$$

$$\widetilde{\mathbf{\Gamma}}_1 = \begin{bmatrix} -0.1637 & 0.0450 & 0.4505 \\ -0.1841 & 0.1687 & 0.1938 \\ -0.0701 & 0.0678 & -0.6178 \end{bmatrix},$$

$$\widetilde{\mathbf{\Gamma}}_2 = \begin{bmatrix} 0.0492 & -0.3409 & 0.5782 \\ 0.0000 & -0.3168 & 0.6822 \\ -0.0844 & 0.0162 & -0.3638 \end{bmatrix},$$

$$\widetilde{\mathbf{\Gamma}}_3 = \begin{bmatrix} 0.0421 & 0.0011 & 0.0352 \\ 0.0244 & 0.1686 & 0.3473 \\ -0.0690 & 0.0082 & -0.2652 \end{bmatrix},$$

and

$$\widetilde{\Sigma}_u = \begin{bmatrix} 0.5659 & 0.0751 & -0.0207 \\ 0.0751 & 0.6165 & 0.0341 \\ -0.0207 & 0.0341 & 0.0654 \end{bmatrix}.$$

Since the three model variables do not have a deterministic linear trend component, one may alternatively absorb the intercept into the error correction

term, as in model (3.1.10). Estimating the model with this restriction of the intercept term, we obtain the ML estimates

$$\widetilde{\alpha} = \begin{bmatrix} 0.2617 & 0.1089 \\ -0.2471 & 0.1192 \\ -0.0352 & 0.0149 \end{bmatrix},$$

$$\widetilde{\beta}^{o\prime} = \begin{bmatrix} -2.1930 & 0.0595 & -0.8534 & 2.1854 \\ -0.3628 & -0.4490 & 1.4726 & 1.6272 \end{bmatrix},$$

$$\widetilde{\mathbf{\Gamma}}_1 = \begin{bmatrix} -0.1639 & 0.0450 & 0.4491 \\ -0.1844 & 0.1688 & 0.1918 \\ -0.0701 & 0.0676 & -0.6171 \end{bmatrix},$$

$$\widetilde{\mathbf{\Gamma}}_2 = \begin{bmatrix} 0.0490 & -0.3409 & 0.5772 \\ -0.0002 & -0.3168 & 0.6807 \\ -0.0843 & 0.0161 & -0.3632 \end{bmatrix},$$

$$\widetilde{\mathbf{\Gamma}}_3 = \begin{bmatrix} 0.0420 & 0.0010 & 0.0347 \\ 0.0242 & 0.1686 & 0.3466 \\ -0.0688 & 0.0081 & -0.2650 \end{bmatrix},$$

and

$$\widetilde{\Sigma}_u = \begin{bmatrix} 0.5658 & 0.0750 & -0.0206 \\ 0.0750 & 0.6165 & 0.0341 \\ -0.0206 & 0.0341 & 0.0655 \end{bmatrix}.$$

Note that these estimates are very similar to those obtained with an unrestricted intercept.

ML Estimation of Gaussian Processes with Known Cointegrating Vectors. A common situation in applied work is that the cointegrating vectors are known. In that case, it makes sense to impose these cointegrating vectors in estimation. For example, the expectations theory of the term structure implies that the spread of interest rates for bonds of different maturities is stationary if the risk premium is stationary. Hence, in a system of two interest rates r_{1t} and r_{2t}, for example, there is a known cointegrating relationship of the form $r_{1t} - r_{2t} = (1, -1)(r_{1t}, r_{2t})'$, allowing ML estimation of α and $\mathbf{\Gamma}$ to condition on $\beta = (1, -1)'$. Similarly, if some nominal variable and the inflation rate are both $I(1)$, whereas the corresponding real variable is $I(0)$, then the nominal variable and inflation are cointegrated with a known cointegration vector. It must be kept in mind, however, that cointegration relations that are implied by economic theory may not be present in the data. Therefore, it is not advisable to blindly impose cointegration vectors without confirming that they are consistent with the data.

Returning to the earlier empirical example, one could postulate that GNP growth is stationary and, hence, constitutes a trivial cointegration relationship, and that the real interest rate is stationary such that $i_t - 4\Delta p_t$ is another cointegration relation, where the factor of 4 arises because the interest rate series is annualized, whereas the inflation rate series is not. In that case, the known cointegration matrix is

$$\beta' = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & -4 \end{bmatrix}$$

which can be used in the ML estimation procedure for ν, α, and $\mathbf{\Gamma}_i$, $i = 1, 2, 3$. The estimates are

$$\widetilde{\nu} = \begin{bmatrix} 0.5156 \\ -0.1828 \\ -0.0593 \end{bmatrix},$$

$$\widetilde{\alpha} = \begin{bmatrix} -0.5304 & -0.0355 \\ 0.4271 & -0.0690 \\ 0.0870 & -0.0067 \end{bmatrix},$$

$$\widetilde{\mathbf{\Gamma}}_1 = \begin{bmatrix} -0.2310 & 0.0259 & 0.3128 \\ -0.1281 & 0.1853 & 0.2636 \\ -0.0808 & 0.0641 & -0.5988 \end{bmatrix},$$

$$\widetilde{\mathbf{\Gamma}}_2 = \begin{bmatrix} 0.0053 & -0.3587 & 0.4930 \\ 0.0364 & -0.3023 & 0.7228 \\ -0.0912 & 0.0136 & -0.3498 \end{bmatrix},$$

$$\widetilde{\mathbf{\Gamma}}_3 = \begin{bmatrix} 0.0221 & -0.0214 & -0.0082 \\ 0.0417 & 0.1844 & 0.3672 \\ -0.0727 & 0.0073 & -0.2574 \end{bmatrix},$$

and

$$\widetilde{\Sigma}_u = \begin{bmatrix} 0.5722 & 0.0680 & -0.0180 \\ 0.0680 & 0.6223 & 0.0330 \\ -0.0180 & 0.0330 & 0.0648 \end{bmatrix}.$$

Note that the Johansen approach uses a purely statistical standardization of the estimates for α and β. Therefore, the ML estimate of β differs substantially from the cointegration matrix we conditioned upon earlier, and so does the estimate of α obtained under the assumption of known and unknown β.

Feasible GLS Estimation. In small samples, the ML estimator occasionally produces estimates far from the true parameter values, as demonstrated in Brüggemann and Lütkepohl (2005), for example. This problem may be alleviated by considering a more robust GLS estimator. Like in the VAR(1) case,

using the normalization $\beta' = [I_r, \beta'_{(K-r)}]$ given in (3.1.5), equation (3.2.8) can be rewritten as

$$R_{0t} - \alpha R_{1t}^{(1)} = \alpha \beta'_{(K-r)} R_{1t}^{(2)} + u_t^*, \tag{3.2.12}$$

where $R_{1t}^{(1)}$ and $R_{1t}^{(2)}$ denote the first r and last $K-r$ components of R_{1t}, respectively. For a given α, the GLS estimator of $\beta'_{(K-r)}$ based on this specification is

$$\widehat{\beta}'_{(K-r)} = (\alpha' \Sigma_u^{-1} \alpha)^{-1} \alpha' \Sigma_u^{-1} \sum_{t=1}^{T} (R_{0t} - \alpha R_{1t}^{(1)}) R_{1t}^{(2)\prime} \left(\sum_{t=1}^{T} R_{1t}^{(2)} R_{1t}^{(2)\prime} \right)^{-1} \tag{3.2.13}$$

(see Lütkepohl 2005, chapter 7). A feasible GLS estimator can be obtained by estimating the matrix $\mathbf{\Pi}$ from $R_{0t} = \mathbf{\Pi} R_{1t} + u_t^*$ with unrestricted equation-by-equation LS, where $\mathbf{\Pi} = [\alpha : \alpha \beta'_{(K-r)}]$. Thus, the first r columns of the estimator for $\mathbf{\Pi}$ can be used as an estimator for α, say $\widehat{\alpha}$, as in the VAR(1) case considered in Section 3.2.1. Substituting this estimator and the corresponding estimator of the white noise covariance matrix, $\widetilde{\Sigma}_u = T^{-1} \sum_{t=1}^{T} \widehat{u}_t^* \widehat{u}_t^{*\prime}$, in expression (3.2.13) yields the feasible GLS estimator

$$\widehat{\widehat{\beta}}'_{(K-r)} = (\widehat{\alpha}' \widetilde{\Sigma}_u^{-1} \widehat{\alpha})^{-1} \widehat{\alpha}' \widetilde{\Sigma}_u^{-1} \sum_{t=1}^{T} (R_{0t} - \widehat{\alpha} R_{1t}^{(1)}) R_{1t}^{(2)\prime} \left(\sum_{t=1}^{T} R_{1t}^{(2)} R_{1t}^{(2)\prime} \right)^{-1}. \tag{3.2.14}$$

This estimator was proposed earlier by Ahn and Reinsel (1990) and Saikkonen (1992) (see also Reinsel 1993, chapter 6). It has the same asymptotic distribution as the ML estimator. Likewise, the asymptotic properties of the associated estimators of α and $\mathbf{\Gamma}$ are the same as in the previous section.

The GLS estimator has two advantages that make it worthwhile to consider this estimator in macroeconomic applications. First, it tends to be much more reliable than the ML estimator in small samples, as illustrated in Brüggemann and Lütkepohl (2005). Second, the GLS estimator can also be adjusted easily to account for conditional heteroskedasticity (see Herwartz and Lütkepohl 2011). Simulation evidence shows reductions in the mean squared error and the mean absolute error of the estimator by more than 30% after allowing for GARCH in the VECM errors.

An alternative approach to estimating cointegration relations was proposed by Engle and Granger (1987) in the early literature on cointegration. Engle and Granger (1987) observed that if there is a single cointegrating relationship between the $I(1)$ variables $y_{1t}, \ldots, y_{Kt}$, the cointegrating relationship can be estimated consistently by regressing y_{1t} on $y_{2t}, \ldots, y_{Kt}$. The GLS method just outlined may be viewed as a generalization of this procedure which allows for serial correlation and possible conditional heteroskedasticity in estimating

the cointegrating parameters. Because the GLS procedure takes into account the full system of equations, it can also be used when there is more than one cointegrating relationship.

Returning to the previously used empirical example, we now illustrate how to estimate the VECM by the feasible GLS estimation method. We first estimate α and Σ_u from $R_{0t} = \mathbf{\Pi} R_{1t} + u_t^*$ using equation-by-equation LS. The estimates are

$$\widehat{\alpha} = \begin{bmatrix} -0.6092 & -0.0344 \\ 0.5064 & -0.0701 \\ 0.0620 & -0.0063 \end{bmatrix},$$

and

$$\widetilde{\Sigma}_u = \begin{bmatrix} 0.5656 & 0.0746 & -0.0201 \\ 0.0746 & 0.6157 & 0.0351 \\ -0.0201 & 0.0351 & 0.0642 \end{bmatrix}.$$

Then we use expression (3.2.14) to compute the estimate of $\beta'_{(K-r)}$,

$$\widehat{\widehat{\beta}}'_{(K-r)} = \begin{bmatrix} 0.2970 \\ -3.8369 \end{bmatrix}, \quad \text{so that} \quad \widehat{\widehat{\beta}}' = \begin{bmatrix} 1 & 0 & 0.2970 \\ 0 & 1 & -3.8369 \end{bmatrix}.$$

Finally, we compute the estimate of $\mathbf{\Gamma} = [\nu, \mathbf{\Gamma}_1, \ldots, \mathbf{\Gamma}_{p-1}]$ as $\widehat{\widehat{\mathbf{\Gamma}}} = \mathbf{\Gamma}(\widehat{\alpha}\widehat{\widehat{\beta}}')$ such that

$$\widehat{\widehat{\nu}} = \begin{bmatrix} 0.7413 \\ -0.3670 \\ -0.0549 \end{bmatrix},$$

$$\widehat{\widehat{\mathbf{\Gamma}}}_1 = \begin{bmatrix} -0.1674 & 0.0445 & 0.4387 \\ -0.1911 & 0.1676 & 0.1679 \\ -0.0621 & 0.0685 & -0.6076 \end{bmatrix},$$

$$\widehat{\widehat{\mathbf{\Gamma}}}_2 = \begin{bmatrix} 0.0465 & -0.3414 & 0.5701 \\ -0.0048 & -0.3180 & 0.6647 \\ -0.0785 & 0.0165 & -0.3558 \end{bmatrix},$$

$$\widehat{\widehat{\mathbf{\Gamma}}}_3 = \begin{bmatrix} 0.0408 & 0.0000 & 0.0310 \\ 0.0221 & 0.1663 & 0.3381 \\ -0.0656 & 0.0087 & -0.2611 \end{bmatrix},$$

and

$$\widehat{\widehat{\Sigma}}_u = \begin{bmatrix} 0.5657 & 0.0748 & -0.0205 \\ 0.0748 & 0.6159 & 0.0345 \\ -0.0205 & 0.0345 & 0.0656 \end{bmatrix},$$

where the latter estimate is computed as

$$\widehat{\widetilde{\Sigma}}_u = \frac{1}{T}\sum_{t=1}^{T}\widehat{\widetilde{u}}_t\widehat{\widetilde{u}}_t',$$

based on the GLS residuals, $\widehat{\widetilde{u}}_t$. This error covariance matrix estimate differs slightly from the first-stage estimate $\widetilde{\Sigma}_u$.

Estimation with Additional Linear Restrictions on the VECM. If restrictions are imposed on the parameters of the VECM representation, the previously discussed estimation methods may be asymptotically inefficient. An alternative is the use of restricted GLS methods which are easy to implement as long as no overidentifying restrictions are imposed on the cointegration matrix. Zero restrictions are the most common constraints for the parameters of these models. Consider, for example, a bivariate system $y_t = (y_{1t}, y_{2t})'$ of two $I(1)$ variables that are cointegrated such that $\beta' y_t \sim I(0)$. If the cointegrating relation appears only in the first equation of the VECM, the loading vector $\alpha = (\alpha_1, 0)'$ has a zero element. Such a restriction can be taken into account in estimation.

If there are no restrictions on the cointegration matrix, the cointegration parameters may be estimated in the first stage by the previously discussed ML or GLS procedures, ignoring exclusion restrictions on the other parameters. Denote this estimator by $\widehat{\beta}$. In the second stage, the remaining parameters may then be estimated from

$$\Delta y_t = \alpha\widehat{\beta}' y_{t-1} + \mathbf{\Gamma}_1\Delta y_{t-1} + \cdots + \mathbf{\Gamma}_{p-1}\Delta y_{t-p+1} + u_t. \tag{3.2.15}$$

Conditional on $\widehat{\beta}$, this is a linear system, and feasible GLS estimation may be used just like for a restricted VAR in levels. For example, if there are linear restrictions of the form

$$\mathrm{vec}[\alpha, \mathbf{\Gamma}] = R\gamma, \tag{3.2.16}$$

then we can rewrite (3.2.15) as

$$\Delta y_t = \mathrm{vec}\left[[\alpha, \mathbf{\Gamma}]\begin{pmatrix}\widehat{\beta}' y_{t-1}\\ X_{t-1}\end{pmatrix}\right] + u_t = [(y_{t-1}'\widehat{\beta}, X_{t-1}')\otimes I_K]R\gamma + u_t.$$

The parameter vector γ can be estimated by feasible GLS. The restricted estimator of α and $\mathbf{\Gamma}$ then is easily obtained from relationship (3.2.16). If a superconsistent estimator $\widehat{\beta}$ is used for the cointegration matrix, the asymptotic properties of the estimators are the same as if β were known.

There are also many situations in which economic theory suggests specific cointegration parameters. As mentioned earlier, the spread of two interest rates of different maturities may be stationary, even if the interest rates are $I(1)$. Knowledge of cointegrating parameters allows us to restrict elements of the cointegrating matrix β. If there are restrictions on β, one may use nonlinear

optimization algorithms to obtain ML estimators of all parameters simultaneously, provided the β parameters are identified. It is also possible to use a two-step procedure that estimates the restricted β matrix first and in a second step conditions on that estimator. When the entire matrix β is known, that matrix can be used directly in the second stage, of course. Technical estimation problems from restrictions on β obviously arise only if some of the elements of β remain unrestricted. Restricted estimation of β is analyzed, for example, in Johansen (1995), Boswijk and Doornik (2004), and Lütkepohl (2005, chapter 7).

Finally, it is worth noting that including deterministic terms in the ML or GLS estimation procedures is straightforward. All that is required is to include these terms in the list of regressors or the cointegration term as appropriate.

3.2.3 Estimation of Levels VAR Models with Integrated Variables

Recovering the VAR Model in Levels. Using the mapping (3.1.4), the parameters of the levels VAR model corresponding to a VECM can be estimated by substituting any of the VECM estimators discussed in the previous subsection. Denoting the resulting estimator of the levels parameters $\boldsymbol{\alpha} = \text{vec}[A_1, \ldots, A_p]$ by $\widehat{\boldsymbol{\alpha}}_{\text{VECM}}$, it can be shown that

$$\sqrt{T}(\widehat{\boldsymbol{\alpha}}_{\text{VECM}} - \boldsymbol{\alpha}) \xrightarrow{d} \mathcal{N}(0, \Sigma_{\boldsymbol{\alpha}}), \tag{3.2.17}$$

if the cointegrating rank $r > 0$ or the lag order $p > 1$. Although this asymptotic result looks like a standard convergence result, there is one important caveat. In this case, the covariance matrix $\Sigma_{\boldsymbol{\alpha}}$ is in general singular. In fact, it is the same covariance matrix that one would obtain if the cointegration matrix β were given and the α and $\mathbf{\Gamma}$ were estimated conditional on β. Since $\widehat{\boldsymbol{\alpha}}$ is an estimator that satisfies all restrictions implied by the VECM, the singularity of the asymptotic distribution is no surprise. As already discussed in Section 3.2.1, one implication of this result is that some of the conventional statistics for inference on the parameters do not have standard asymptotic distributions. In particular, Wald statistics for linear parameter restrictions may not have their usual asymptotic χ^2 distributions under the null hypothesis. In fact, if the VAR order is $p = 1$, even t-statistics may not be asymptotically standard normal anymore, as shown in Section 3.2.1.

Estimation of the VAR Model in Levels. In practice, the cointegration structure is often unknown. At best it can be estimated and is subject to estimation uncertainty. In that case, an alternative approach is to estimate the VAR in levels without imposing cointegration restrictions. As in the VAR(1) model considered in Section 3.2.1, the LS estimator of this model, denoted by $\widehat{\boldsymbol{\alpha}}_{\text{LS}}$, has

exactly the same asymptotic distribution as the $\widehat{\boldsymbol{\alpha}}_{\text{VECM}}$ estimator in (3.2.17),

$$\sqrt{T}(\widehat{\boldsymbol{\alpha}}_{\text{LS}} - \boldsymbol{\alpha}) \xrightarrow{d} \mathcal{N}(0, \Sigma_{\boldsymbol{\alpha}}), \tag{3.2.18}$$

even if the cointegration restrictions are not imposed in estimation. As before, the reason is that the cointegration parameters and, hence, the cointegrating relationships are estimated superconsistently. Also, as before, the common trends in some components of the process for y_t induce a singularity in the asymptotic covariance matrix. Not knowing the precise structure of $\Sigma_{\boldsymbol{\alpha}}$ is a problem in general because it implies that the distributions of some test statistics are unknown.

As discussed in Chapter 2, a possible remedy for this problem is to add a redundant lag to the VAR in levels and to estimate a VAR($p + 1$) model instead of a VAR(p) model. If the highest order of integration of the variables is $I(1)$, then this lag augmentation approach ensures that the parameter matrices associated with the first p lags have a nonsingular asymptotic distribution. Thus, tests for hypotheses involving only parameters from the first p slope parameter matrices retain their standard asymptotic properties. Because we know that A_{p+1} is zero by construction, there is no need to test hypotheses about that term.

More generally, if some of the components are $I(d)$ and none of the component series has a higher order of integration, fitting a VAR($p + d$) solves the singularity problem for the LS estimator of the parameters associated with the first p lags (Toda and Yamamoto 1995; Dolado and Lütkepohl 1996). Hence, lag augmentation or overfitting can be used more generally to overcome problems with asymptotic distributions. Of course, using this device has a cost in terms of lower estimation precision.

Sieve Estimation. As dicussed in Chapter 2, fitting a finite-order VAR model can be justified as an approximation to an infinite-order VAR process. The asymptotic theory justifying such sieve approximations assumes that the lag order increases with the sample size at a suitable rate. The same device can be used for VECMs. Using the framework of Saikkonen (1992), Saikkonen and Lütkepohl (1996) state conditions that ensure the same asymptotic properties of the GLS estimator for the cointegrating matrix β as in the finite-order case. They also show that under their conditions, the VECM parameters, after suitable rescaling, converge to a Gaussian marginal limiting distribution, as in the case of a true finite-order VECM. Thus, even if the finite-order VECM used for a particular system of variables is just an approximation, standard methods of estimation and inference for VECM parameters remain valid in this case.

It should be noted that Saikkonen and Lütkepohl's results are derived for processes without deterministic terms. These results can be extended to the

case of an intercept term capturing a nonzero mean of the process. Extensions for other deterministic terms appear to be nontrivial.

Moreover, all these results presume that the cointegrating vectors have been correctly specified and that the variables are either $I(0)$ or $I(1)$. If the model variables are not well approximated by either the $I(0)$ or $I(1)$ assumption, yet another option is to appeal to local-to-unity asymptotics (also known as near unit root asymptotics) or to treat the variables as fractionally integrated.

Local-to-Unity Asymptotics. In general, small-sample distributions of estimators of VAR models in levels may not be well approximated by their asymptotic counterparts. For VAR(1) models, the small-sample distortions were found to be particularly troublesome. In response to this problem, Stock (1996), Phillips (1998), and Pesavento and Rossi (2006), among others, have considered alternative asymptotic approximations based on models in which some roots of the autoregressive polynomial are close to 1 but not exactly equal to 1. We discuss that approach next.

A possibly integrated or cointegrated VAR(p) model without deterministic terms can alternatively be written as

$$y_t = Cy_{t-1} + \mathbf{\Gamma}_1 \Delta y_{t-1} + \cdots + \mathbf{\Gamma}_{p-1} \Delta y_{t-p+1} + u_t, \tag{3.2.19}$$

where

$$C = I_K + \alpha\beta' = I_K + \mathbf{\Pi} = \sum_{i=1}^{p} A_i.$$

Thus, if the cointegrating rank is zero, $C = I_K$ and, more generally, if $\mathbf{\Pi}$ is close to zero, the $K \times K$ matrix C is close to an identity matrix.

Local-to-unity asymptotics are motivated by the observation that in small samples one cannot reliably discriminate between the hypothesis that C equals I_K and the hypothesis that it does not. Following Stock (1996), Phillips (1998), and Pesavento and Rossi (2006), among others, this situation may be modeled by postulating that $C = I_K + \Lambda/T$ in population where Λ is a $K \times K$ diagonal matrix with fixed negative elements $\lambda_1, \lambda_2, \ldots, \lambda_K$ along the main diagonal. For any finite T, the diagonal elements of C are smaller than 1. This setup differs from the standard asymptotic thought experiment, in which the parameter matrix C is fixed as $T \to \infty$. Instead, C is treated as local-to-unity or near unity in the sense that C becomes arbitrarily close to I_K, as $T \to \infty$.

Modeling one or more roots in the vector autoregressive lag polynomial as local to unity is only a statistical device for obtaining better approximations to the finite-sample distribution of the estimator of interest. It does not mean that we believe that the autoregressive roots of the DGP actually depend on the observed sample size. The fact that we allow the coefficient matrix C to depend on the sample size ensures that it does not become easier with increasing T to

tell the difference between the diagonal elements of C being exactly unity or below unity, because the target to be estimated shifts closer to I_K at the same rate as the estimator of C approaches the target. In other words, we describe a situation in which it is not possible to tell whether the roots are unity or not, even asymptotically.[3]

A key difference from conventional asymptotics is that, in the local-to-unity framework, C cannot be estimated consistently and that the asymptotic distribution of many estimators of interest in applied work is no longer normal, but nonstandard. Methods of constructing confidence intervals based on such nonstandard asymptotic approximations are discussed in Chapter 12. Whether local-to-unity asymptotics generate more accurate approximations to the finite-sample distribution of an estimator than do conventional asymptotic thought experiments is an empirical question. The fact that an observed time series is highly persistent does not automatically imply that it must be modeled using the local-to-unity framework. As we have seen, inference based on the level VAR(p) model (or based on the first p coefficients of a VAR($p+1$) model) may also be considered and indeed may be more robust to possible (near) cointegration of unknown form among the model variables. Ultimately, the choice between these asymptotic approximations is determined by their finite-sample accuracy.

Fractional Integration. An alternative approach to modeling persistence in the variables that may not be captured well by unit roots, is to treat the variables as fractionally integrated. A general representation of a system of fractionally integrated and possibly cointegrated (or cofractional) variables is

$$\Delta^d y_t = \Delta^{d-b} L_b \alpha \beta' y_t + \sum_{i=1}^{p} \mathbf{\Gamma}_i \Delta^d L_b^i y_t + u_t, \tag{3.2.20}$$

where for real numbers d, Δ^d signifies the fractional differencing operator defined in Chapter 2 as

$$\Delta^d = (1-L)^d = \sum_{i=0}^{\infty} (-1)^i \binom{d}{i} L^i$$

and L_b is the fractional lag operator defined as

$$L_b = 1 - \Delta^b$$

(see Johansen 2008; Johansen and Nielsen 2012). The model (3.2.20) assumes variables with fractional integration order d and cofractional order $d - b \geq 0$. The $K \times r$ matrices α and β are assumed to have rank r and $\beta' y_t$ is

[3] Setting the off-diagonal elements of Λ to zero rules out that any of the variables is $I(2)$ (see Elliott 1998; Phillips 1998).

fractionally integrated of order $d - b$. For $b > 0$, variables with this property are called cofractional of order $d - b$. For given r, under the assumption that $u_t \overset{iid}{\sim} \mathcal{N}(0, \Sigma_u)$, the parameters of the model, including the fractional parameters b and d, can be estimated by ML. The asymptotic theory is developed by Johansen and Nielsen (2012).

The use of fractionally integrated and/or fractionally cointegrated models in structural VAR analysis has been very limited thus far. In practice, it is common to explore the fractional integration orders of the model variables first, before estimating the model conditional on these orders (see, e.g., Tschernig, Weber, and Weigand 2013). How well this approach works in small samples is not known. Nor is it known whether this approach leads to a better understanding of the DGP than, for example, using a local-to-unity approach. Which of these approaches is preferable in practice is likely to depend on the specific DGP. In either case, departing from the standard $I(1)$ setup complicates the analysis. Tschernig, Weber, and Weigand (2013) present simulation evidence that suggests that such complications may be worth contemplating when the deviations from the $I(1)$ model are sufficiently large and large samples are available.

3.3 Model Specification

Specifying VECMs involves choosing the lag order and determining the cointgrating rank. These two issues are discussed next.

3.3.1 Choosing the Lag Order

As in the stationary case, the VAR order can be chosen by sequential tests or model selection criteria if there are $I(1)$ variables among the components of the VAR process. If some of the variables are $I(1)$, the usual LR or Wald tests for the lag order have standard asymptotic χ^2 distributions under the null hypothesis as long as one does not test the hypothesis $\mathbb{H}_0 : A_1 = 0$.

Tests for residual autocorrelation can again be used in a bottom-up sequential testing strategy for lag-order selection as in Chapter 2. However, if there are integrated variables, the approximate distribution of the Portmanteau tests must be adjusted, as pointed out by Brüggemann, Lütkepohl, and Saikkonen (2006). For cointegrated processes the degrees of freedom also depend on the cointegrating rank. The approximate distribution for a VECM with cointegrating rank r and $p - 1$ lagged differences on the right-hand side is $\chi^2(K^2h - K^2(p-1) - Kr)$. Since the cointegrating rank is typically unknown at the time of lag-order selection, the Portmanteau test is not useful for lag-order selection. In contrast, the Breusch-Godfrey LM test for residual autocorrelation can be applied to levels VAR processes with unknown cointegrating rank. Its asymptotic $\chi^2(hK^2)$ distribution under the null hypothesis is valid for

both $I(0)$ and $I(1)$ systems, as shown in Brüggemann, Lütkepohl, and Saikkonen (2006). Moreover, the information criteria discussed in Section 2.4 of the previous chapter maintain their asymptotic properties and, hence, can be used for cointegrated processes as well with the same justification as for stationary processes (see Paulsen 1984).

3.3.2 Specifying the Cointegrating Rank

Several proposals have been made for determining the cointegrating rank of a VAR process. Many of them are reviewed and compared in Hubrich, Lütkepohl, and Saikkonen (2001). Generally, a good case can be made for using the Johansen (1995) likelihood ratio approach (and its modifications) of testing for the cointegrating rank under the maintained assumption of Gaussian errors. Even if the DGP is not Gaussian, the resulting pseudo-LR tests have better properties than do many competitors. These tests are also attractive from a computational point of view because, for a given cointegrating rank r, ML estimates and, hence, the maximum of the likelihood function are easy to compute (see Section 3.2.2). Of course, in some cases alternative tests may be more suitable, as discussed in Hubrich, Lütkepohl, and Saikkonen (2001).

Denoting the matrix $\alpha\beta'$ in the error correction term by $\mathbf{\Pi}$, as before, the following sequence of hypotheses may be considered for selecting the cointegrating rank:

$$\mathbb{H}_0(r_0) : \operatorname{rank}(\mathbf{\Pi}) = r_0 \text{ versus } \mathbb{H}_1(r_0) : \operatorname{rank}(\mathbf{\Pi}) > r_0, \quad r_0 = 0, \ldots, K-1. \tag{3.3.1}$$

The corresponding LR test statistic is

$$LR^{trace}(r_0) = -2(\log l(r_0) - \log l(K)) = -T \sum_{i=r_0+1}^{K} \log(1-\lambda_i),$$

where $l(r)$ denotes the maximum of the Gaussian likelihood function (3.2.9) given the cointegration rank r and λ_i is the i^{th} eigenvalue obtained by the Johansen procedure in Section 3.2.2. This test is often referred to as the trace test. The cointegrating rank specified in the first null hypothesis that cannot be rejected is chosen as the estimate for the true cointegrating rank r. If $\mathbb{H}_0(0)$, the first null hypothesis in this sequence, cannot be rejected, we proceed with a VAR process in first differences. If all the null hypotheses are rejected including $\mathbb{H}_0(K-1)$, the process is treated as $I(0)$, and a levels VAR model is specified.

Instead of testing the sequence of hypotheses specified in (3.3.1), one may alternatively test the sequence

$$\mathbb{H}_0(r_0) : \operatorname{rank}(\mathbf{\Pi}) = r_0 \text{ versus } \mathbb{H}_a(r_0+1) : \operatorname{rank}(\mathbf{\Pi}) = r_0 + 1, \quad r_0 = 0, \ldots, K-1. \tag{3.3.2}$$

Table 3.1. *Trace and Maximum Eigenvalue Tests for Cointegrating Rank*

λ_i	$\mathbb{H}_0$	LR^{trace}	critical value	LR^{max}	critical value
		unrestricted intercept			
0.2020	$r = 0$	60.6041	28.71	47.1489	18.90
0.0398	$r = 1$	13.4552	15.66	8.4972	12.91
0.0234	$r = 2$	4.9580	6.50	4.9580	6.50
		restricted intercept			
0.2020	$r = 0$	60.6968	32.00	47.1550	19.77
0.0399	$r = 1$	13.5418	17.85	8.5117	13.75
0.0238	$r = 2$	5.0301	7.52	5.0301	7.52

Note: The critical values for the upper panel are taken from table 1.1* and the critical values for the lower panel from table 1* of Osterwald-Lenum (1992) for a 10% level test.

In other words, a specific rank r_0 is tested against the rank $r_0 + 1$. The LR test statistic for this pair of hypotheses is

$$LR^{max}(r_0) = -T \log(1 - \lambda_{r_0+1}).$$

This test is known as the maximum eigenvalue test. As in the case of the trace test, the test sequence terminates when the null cannot be rejected.

The LR test statistics corresponding to the null hypotheses in (3.3.1) and (3.3.2) have nonstandard asymptotic distributions. Their asymptotic distributions depend on the difference $K - r_0$ and on the deterministic terms included in the DGP, but they do not depend on the short-term dynamics. More precisely, under $\mathbb{H}_0$ the asymptotic distribution of the LR test corresponding to (3.3.1) is the trace of a matrix functional of multivariate Brownian motions, and that corresponding to (3.3.2) is the maximum eigenvalue of the corresponding matrix functional, which explains the specific names of the tests as trace and maximum eigenvalue tests. Critical values for various possible sets of deterministic components such as constants and linear trends have been computed by simulation methods and are available in the literature (e.g., Johansen 1995, chapter 15; Osterwald-Lenum 1992).

An Empirical Illustration. To illustrate the use of cointegration rank tests, we again utilize the trivariate empirical example from Section 3.2.2. Table 3.1 shows the eigenvalues obtained for a model with unrestricted intercept and for a model with the intercept contained only in the cointegration relations. The table also displays the corresponding values of the trace and maximum eigenvalue test statistics together with critical values for a test level of 10%. The level refers to each individual test and not to the joint level of the testing sequence.

Clearly, all tests reject the cointegrating rank $r = 0$, but they do not reject $r = 1$ and $r = 2$. Thus, based on Table 3.1, one may conclude that the preferred

cointegrating rank is $r = 1$. Because these tests tend to have low power in small samples, however, using the model with $r = 2$ rather than a model with $r = 1$ may also be justified when economic considerations suggest two cointegration vectors.

Size and Power Considerations. As in the case of sequential tests for the lag order, the overall size of tests for the cointegrating rank differs from the nominal size chosen for individual tests in the sequence. It is not clear how to control the overall size of these tests. The small-sample size and power of the trace and maximum eigenvalue tests are compared in Cheung and Lai (1993) and Toda (1994, 1995), among others. The performance of these tests is often similar. There is no clear ranking between the two types of testing sequences. Generally, the power of both types of tests tends to be low in many situations of practical interest.

The power of these tests may be improved by specifying the deterministic terms as tightly as possible. For example, if there is no deterministic linear trend term, it is desirable to perform the cointegrating rank tests without such terms. On the other hand, incorrectly omitting them may cause major size distortions. The asymptotic theory for testing hypotheses regarding the deterministic terms provided by Johansen (1995) can be helpful in this respect.

If a linear trend is required, but it is unknown whether or not this trend is orthogonal to the cointegration relationships, Demetrescu, Lütkepohl, and Saikkonen (2009) propose to apply rank tests with both alternative trend terms and to reject the null hypothesis if one of these tests rejects. They provide an asymptotic justification for this procedure.

Tests for the cointegrating rank have also been developed for the case of a structural break in the deterministic term either in the form of a level shift or a break in the trend slope, or both. In this case the critical values of the LR tests also depend on the timing of the break. This feature is inconvenient if the break point is not known a priori and has to be estimated. In that case, a test variant proposed by Saikkonen and Lütkepohl (2000a, 2000b) may be preferable. They suggest to estimate the deterministic term first by a GLS procedure and to adjust the data before applying a modified LR-type test to the adjusted system. The advantage is that the asymptotic null distribution of the test statistic does not depend on the break point, if only a level shift is considered. This fact facilitates the development of procedures that work even when the break date is unknown (e.g., Lütkepohl, Saikkonen, and Trenkler 2004; Saikkonen, Lütkepohl, and Trenkler 2006).

Although the short-run dynamics do not matter for the asymptotic properties of the rank test, they have a substantial impact in small and moderate samples. Therefore, the choice of the lag order p is quite important in conducting rank tests. On the one hand, choosing p rather large to avoid misspecifying the short-run dynamics tends to cause a substantial loss in the power of the

cointegrating rank tests. On the other hand, choosing the lag order too small may lead to dramatic size distortions. In a small-sample simulation study, Lütkepohl and Saikkonen (1999) conclude that using the AIC criterion for lag-order selection is a good compromise when determining the cointegrating rank.

There are many other proposals for modifying and improving the Johansen approach to cointegration testing. For example, Johansen (2002) presents a Bartlett correction designed to improve the performance of the Johansen cointegration tests in small samples. For further discussion of this and other approaches the reader is referred to Hubrich, Lütkepohl, and Saikkonen (2001). At present it appears that the Johansen approach should be the default among tests for the cointegrating rank, unless there is a compelling reason for using another type of test.

Subsystem Tests. Clearly, the Johansen approach to testing for the cointegrating rank has its drawbacks, in particular when used in large-dimensional systems or when many lags are necessary to capture the short-term dynamics. In this case, the test may lack the power to detect all cointegration relationships and, as a result, may understate the true cointegrating rank (see Gonzalo and Pitarakis 1999). Hence, it is recommended to apply cointegration tests to all possible subsystems as well and to verify whether the results are consistent with those for the full model. For example, in a K-dimensional system where all variables are individually $I(1)$, if all variables are cointegrated in pairs, the cointegrating rank must be $K - 1$. Cointegration for the bivariate subsystems may be easier to analyze than cointegration in the full K-dimensional system. This observation suggests that one should analyze the subsystems first and then assess whether the subsystem results are consistent with the results for the full system, taking into account that the cointegrating rank tests may have reduced power for larger systems.

As an explicit example, consider a four-dimensional system $y_t = (y_{1t}, y_{2t}, y_{3t}, y_{4t})'$ and suppose that the first three components are individually $I(1)$. Put differently, $y_{it} \sim I(1)$ for $i = 1, 2, 3$, and $y_{4t} \sim I(0)$. If there are two linearly independent cointegrating relations between the first three variables, it is easy to see that all pairs $(y_{1t}, y_{2t})'$, $(y_{2t}, y_{3t})'$, and $(y_{1t}, y_{3t})'$ are also cointegrated. Moreover, taking into account that y_{4t} is $I(0)$, the cointegrating rank of y_t is three. If unit root tests are consistent with y_{it} being $I(1)$ for $i = 1, 2, 3$, and $y_{4t} \sim I(0)$, then we can proceed to testing cointegration in the three pairs $(y_{1t}, y_{2t})'$, $(y_{2t}, y_{3t})'$, and $(y_{1t}, y_{3t})'$. If the tests are consistent with cointegration in all three pairs, and if we also take into account that y_{4t} is $I(0)$, we can conclude that y_t has cointegrating rank three. Given that there are error probabilities associated with all these tests, we may want to follow up with a sequential test for the cointegrating rank of y_t applied to the full system. It is quite possible that in this case the testing sequence rejects ranks zero and 1 but

not rank two. Had we only applied the testing sequence to the full system, we might have incorrectly concluded that the cointegrating rank is two, whereas taking into account the results from the previous subsystem tests, we conclude that the rank is three and that the non-rejection of rank two for the full system is just due to a lack of power against the alternative of a larger rank.

In this example, other seemingly conflicting test results are, of course, possible. For example, the tests may suggest cointegration between y_{1t} and y_{2t} as well as y_{2t} and y_{3t}. In that case, y_{1t} and y_{3t} are necessarily also cointegrated. Yet, a cointegration rank test may not reject rank zero for $(y_{1t}, y_{3t})'$. Again, that result could be due to a lack of power. It must be kept in mind that not rejecting a null hypothesis does not establish the validity of the null hypothesis, but only that there is not enough sample information to be sure beyond a reasonable doubt that the null is false. If looking at the sample information from some other angle makes the data speak more clearly, then there is nothing wrong with relying on that information. Hence, in that example working with a cointegrating rank of $r = 2$ for the three-dimensional subsystem $(y_{1t}, y_{2t}, y_{3t})'$ and with $r = 3$ for the full four-dimensional system would be a sensible choice.

3.4 Diagnostic Tests

Most of the diagnostic tools discussed in the previous chapter for VAR models estimated in levels are also applicable to cointegrated VAR models and VECMs. We already mentioned tests for residual autocorrelation. Although Portmanteau tests are not suitable for models with unknown cointegrating rank because their distribution depends on r, they can be used to test for autocorrelation in the innovations of a given VECM that is assumed to be valid under the null hypothesis. In that case, the cointegrating rank is assumed to be correctly specified, and hence the Portmanteau test has an approximate $\chi^2(K^2h - K^2(p-1) - Kr)$ distribution. As mentioned earlier, no adjustments are necessary for the Breusch-Godfrey LM test for error autocorrelation in a VECM or in a VAR model with integrated variables (see Brüggemann, Lütkepohl, and Saikkonen 2006).

The tests for nonnormality mentioned in the previous chapter can also be applied to the residuals of a VECM. The asymptotic distributions of the test statistics are not affected by $I(1)$ variables in the model. This result follows from the superconsistency of the estimator for the cointegration matrix and the properties of the empirical moment matrices of the integrated model variables (see Kilian and Demiroglu 2000).

It is also straightforward to extend tests for structural change such as the Chow test to the case of cointegrated processes. Suppose that a change in the parameters of the VECM (3.2.6) is suspected after period $T_1 < T$. In the context of VECMs one may then be interested in testing

$$\mathbb{H}_0 : \beta_{(1)} = \beta_{(2)}, \alpha_{(1)} = \alpha_{(2)}, \mathbf{\Gamma}_{(1)} = \mathbf{\Gamma}_{(2)}, \tag{3.4.1}$$

where the subscripts (1) and (2) refer to the first and second subperiods, respectively, against the alternative that at least one of the equalities is violated. The relevant Wald and LR tests have asymptotic χ^2 distributions. The degrees of freedom have to account for the fact that a nonsingular asymptotic distribution for the estimator of β is only obtained upon suitable normalization. Thus, $\beta_{(1)} = \beta_{(2)}$ implies only $r(K - r)$ restrictions. The corresponding LR statistic for testing the null hypothesis (3.4.1) hence has a limiting χ^2 distribution with $r(K - r) + rK + (p - 1)K^2$ degrees of freedom. Tests for constancy of only a subset of the parameters can be constructed analogously (see Hansen 2003). Of course, these tests can be extended to models with deterministic terms.

3.5 The Benefits of the VECM Representation

Ultimately, there are two reasons for users of structural VAR models to be interested in the VECM representation. One reason is the efficiency gains in estimating the reduced-form VAR model when the VECM is correctly specified. Once the VECM has been estimated, however, it is often convenient to represent the estimates as a VAR model in levels, as shown in Section 3.1. This representation facilitates the construction of forecasts and impulse responses, in particular. Expressing forecasts and impulse responses instead in terms of the VECM parameters does not provide any new insights. Similarly, imposing known unit roots and imposing known cointegration restrictions on the VAR model may improve the power of statistical tests such as Granger causality tests (see Lütkepohl and Reimers 1992b).

The other reason for considering VECMs is that they facilitate the imposition of restrictions on the long-run effects of structural shocks in the VAR model, which extends the range of identifying assumptions used for structural impulse response analysis. This point is discussed in Chapter 10.

3.6 Practical Issues

In practice, it is rarely clear when to use the VECM framework as opposed to a VAR in levels or in differences. Obviously, if we were sure of the existence of a unit root, we ought to impose it in estimation. If we are not sure about the presence of a unit root, as is typically the case in practice, the situation changes. On the one hand, incorrectly imposing a unit root results in overdifferencing of the data, rendering the VAR estimator inconsistent under standard assumptions. Failing to impose a unit root when the unit root is correct, on the other hand, preserves consistency. It only causes a reduction in the precision of the LS estimator and worsens its small-sample bias. Thus, the consequences of correctly imposing a unit root and of incorrectly imposing a unit root are asymmetric. This conclusion also extends to the question of whether to impose the cointegrating rank or the cointegrating vector in estimation. One practical strategy used in the literature is to rely on VAR models in levels that tend to be

robust to alternative specifications of the cointegrating rank and vectors. The potential cost of this approach, of course, is that we may not exploit all the economic structure in the DGP. In addition, the use of a level representation prevents us from using certain types of identification schemes for structural shocks that have been popular in the literature (see Chapter 10). An alternative strategy used in the literature is to specify a VECM if that model can be economically motivated and if the data do not object to this specification. This approach requires verifying that unit root and cointegration tests do not contradict the properties of the VECM. It also requires establishing that other diagnostic tests do not raise concerns about the model specification and showing that key features of the VECM are robust to relaxing the VECM specification. This strategy is not without risks, however, as discussed in the next section.

3.6.1 Limitations of Tests for Unit Roots and Cointegration

It may seem that the question of whether the model should be specified as a VAR model in levels, as a VAR model in differences, or as a VECM could be resolved by implementing a battery of pretests for unit roots and cointegration. This is not the case. There are two distinct problems. First, these tests cannot be used to confirm the features specified under the null hypothesis. For example, many empirical studies rely on evidence that unit root tests fail to reject the null hypothesis of a unit root as justification for imposing the unit root in estimation. However, a non-rejection of the null hypothesis only means that there is insufficient evidence to rule out unit roots beyond a reasonable doubt. This outcome may arise because the unit root is true or because the test lacks power against the alternative. All we can say is that the data are consistent with a unit root, just as they are consistent with the absence of a unit root. Much the same type of problem afflicts commonly used tests of the $I(0)$ null hypothesis. Similar problems of interpretation also arise in tests for cointegration.

Second, not only do these pretests suffer from low power, but Elliott and Stock (1994) and Cavanagh, Elliott, and Stock (1995), among others, demonstrate that the use of unit root pretests is invalid in environments when the dominant root is local-to-unity. Second-stage inference based on these pretests for unit roots exhibits substantial size distortions in empirically plausible situations. The same concern applies to pretests for the cointegration rank (see Elliott 1998). Although the local-to-unity model merely represents a thought experiment that may or may not approximate the DGP well, these studies show that inference drawn after doing pretests for unit roots and cointegration can be very misleading even for DGPs close to a VECM.

3.6.2 Alternative Approaches

There are two main alternatives in practice. One approach is to make explicit that the author wishes to impose some unit roots or some cointegration

relationship in estimation, while recognizing that such an assumption could potentially invalidate the empirical results. One may also examine the sensitivity of the results to alternative modeling assumptions. Of course, there is no guarantee that the results will be quantitatively or qualitatively consistent across specifications. The other approach is to specify the VAR model in levels. The use of the levels specification not only avoids the unit root issue. It also avoids the controversial issue of which cointegration restrictions to impose in estimation. It is useful to keep in mind, however, that the levels VAR model as well is only an approximation to the DGP.

Specifying the VAR model in levels is not without drawbacks either. First, this specification cannot be used for imposing restrictions on the long-run behavior of the data, as discussed in Chapter 10. Second, there is the question of how to conduct inference in that case. One option is to rely on the alternative asymptotic theory for near-unit root processes. Existing results within this theoretical framework, however, are limited to impulse responses, and the implementation of these methods of inference can be computationally challenging in some cases. Another option is to rely on the Bayesian methods of estimation and inference discusssed in Chapter 5. The latter approach is not without its own challenges, one of which is the specification of the prior. The key difference is that standard Bayesian distribution theory is invariant to the knife-edge case of an exact unit root, which simplifies the analysis, whereas classical distribution theory is discontinuous at the unit circle. Precisely because the data are fairly uninformative about the presence of unit roots and cointegration, however, Bayesian estimates are sensitive to prior information about the long-run behavior of the data, leaving the unit root question unresolved. The third option is to exploit the fact that for higher-order autoregressive models in some cases standard methods of inference remain asymptotically valid even in the possible presence of unit roots and cointegration, as discussed in Section 3.2.3 (see Sims, Stock, and Watson 1990). Dealing with the singularities in the asymptotic covariance of many estimators may require the use of lag-augmented VAR models, however (see, e.g., Dolado and Lütkepohl 1996).

Even when the estimator of the VAR model in levels remains asymptotically valid in the presence of possible unit roots and cointegration, its small-sample properties may be unsatisfactory. The main concern in practice is the higher small-sample bias in the levels specification compared with models that impose unit roots. How severe this small-sample bias is depends on the model and on the sample size. This bias problem is greatly exacerbated by the inclusion of a deterministic time trend. A Monte Carlo study in Inoue and Kilian (2002a) demonstrates that even in univariate AR(2) models, the first-order approximation for the slope parameter in the levels model with deterministic time trend remains poor until the sample size exceeds $T = 500$. In contrast, without the deterministic time trend, $T = 300$ is perfectly adequate.

In practice, few researchers in empirical macroeconomics include a deterministic time trend in the VAR model. One reason is that many macroeconomic time series are highly persistent, and a deterministic time trend in conjunction with a (near) unit root in the autoregressive lag polynomial implies a (near) quadratic trend in y_t. This specification seems implausible when the trend in the data appears linear, as is the case for standard macroeconomic aggregates such as U.S. industrial production or real GDP. Moreover, the inclusion of a time trend in such models invites overfitting in small samples. This realization prompted Sims (1987) to argue for the exclusion of the deterministic trend term. This argument is uncontroversial from a Bayesian point of view. From a frequentist perspective, it requires taking a stand on the trend model in favor of a (near) stochastic trend. This is effectively what many non-Bayesian users of VAR models have done when adopting the same levels specification as Sims, although this point is rarely made explicit. While this additional restriction is immaterial in modeling nontrending series such as interest rates, for example, it is not innocuous when dealing with trending data such as real GDP or industrial production. The exclusion of the deterministic time trend effectively means that nonstochastic trending behavior in the data can only arise from a drift term in the VAR model. In this sense, the unrestricted VAR model in levels nests $I(1)$ processes and local-to-unity processes. Even if we are willing to grant the existence of unit roots, however, this approach is still useful in that it frees the user from the need to commit to a specific cointegration structure in estimating the VAR model.

Whether one prefers the VECM approach or the VAR in levels in the end comes down to how certain one is of the implied restrictions. Although the alternative of working with the VAR in levels is not without its own drawbacks, tentative simulation evidence in Gospodinov, Herrera, and Pesavento (2013) suggests that typically estimates based on the VAR model in levels are more accurate than estimates from models selected based on pretests. The relevance of such simulation results for applied work, of course, depends on how closely the assumed underlying DGP matches the features of the actual DGP.