Chapter 16

VECTOR AUTOREGRESSIONS WITH UNIT ROOT NONSTATIONARY PROCESSES

This chapter explains econometric methods related to VARs and cointegration. We first introduce a broader concept of cointegration that allows us to treat the case in which a vector time series includes both stationary and nonstationary variables. In the previous chapters, cointegration is only defined for a vector time series that does not include stationary variables. Then we discuss a method to impose long-run restrictions for VARs with stationary variables for which the nonstationary variables in the vector time series are not cointegrated. We will explain various representations of a cointegrated system such as Vector Error Correction Model (VECM) and Phillips' triangular representation. Then we will present methods to impose long-run restrictions imposed on Phillips' triangular representation and VECM representation. We will introduce a structural Error Correction Model (ECM) by considering a foreign exchange rate model in which prices and the exchange rate adjusts toward a long-run equilibrium level. A method to estimate the structural speed of the adjustment coefficient toward the long-run equilibrium level will be discussed. In the Appendix, we will discuss long-run and short-run restrictions imposed on VECM.

16.1 Identification on Structural VAR Models

16.1.1 Long-Run Restrictions for Structural VAR Models

Blanchard and Quah (1989) propose using long-run restrictions to identify the underlying shocks in a VAR system. Let y_t be the logarithm of GDP and u_t be the level of the unemployment rate. Here y_t is assumed to be difference stationary and u_t is assumed to be stationary. Let $\mathbf{y}_t = (\Delta y_t, u_t)'$, and let $\mathbf{e}_t = (e_t^s, e_t^d)'$ be the underlying shocks of the economy, where e_t^d is the demand shock, and e_t^s is the supply shock. It is assumed that the demand and supply shocks are uncorrelated, and that $\mathbf{y}_t$ has an MA representation in terms of $\mathbf{e}_t$:

$$
\begin{aligned}
\text{(16.1)} \qquad \mathbf{y}_t &= \boldsymbol{\mu} + \boldsymbol{\Phi}(L)\mathbf{e}_t \\
&= \boldsymbol{\mu} + \boldsymbol{\Phi}_0\mathbf{e}_t + \boldsymbol{\Phi}_1\mathbf{e}_{t-1} + \boldsymbol{\Phi}_2\mathbf{e}_{t-2} + \cdots,
\end{aligned}
$$

where $\boldsymbol{\Phi}(1)$ is normalized so that its principal diagonal components are 1's, and $E(\mathbf{e}_t\mathbf{e}_t') = \boldsymbol{\Lambda}$.

The long-run restrictions are that the demand shock does not have any long-run effect, and the supply shock does not have any long-run effect on unemployment, but may have a long-run effect on the level of output. These restrictions imply that the matrix $\boldsymbol{\Phi}(1)$ is lower triangular.

Let $\mathbf{y}_t = \boldsymbol{\mu} + \boldsymbol{\Psi}(L)\boldsymbol{\epsilon}_t$ be the Wold representation, which can be estimated by inverting the VAR representation for $\mathbf{y}_t$. Then $\boldsymbol{\epsilon}_t = \boldsymbol{\Phi}_0\mathbf{e}_t$, $\boldsymbol{\Sigma}_\epsilon = E(\boldsymbol{\epsilon}_t\boldsymbol{\epsilon}_t') = \boldsymbol{\Phi}_0\boldsymbol{\Lambda}\boldsymbol{\Phi}_0'$, and $\boldsymbol{\Phi}_j = \boldsymbol{\Psi}_j\boldsymbol{\Phi}_0$ for all j. Once $\boldsymbol{\Phi}_0$ is known, we can obtain $\mathbf{e}_t$ from $\boldsymbol{\epsilon}_t$, and $\boldsymbol{\Phi}_j$ from $\boldsymbol{\Psi}_j$. Is $\boldsymbol{\Phi}_0$ identified? An informal argument by Blanchard and Quah suggest that it is. Given $\mathbf{\Sigma}_\epsilon$, the equation $\mathbf{\Phi}_0\mathbf{\Lambda}\mathbf{\Phi}_0' = \mathbf{\Sigma}_\epsilon$ gives three restrictions because $\mathbf{\Sigma}_\epsilon$ is symmetric. Given $\mathbf{\Psi}(1)$, the equation that the upper right-hand entry of $\mathbf{\Phi}(1)$ is zero gives one more restriction. There exist four restrictions for four unknown parameters in $\mathbf{\Phi}_0$.

The assumption that $\mathbf{\Phi}(1)$ is lower triangular gives $\frac{n(n-1)}{2}$ necessary conditions. From $\mathbf{\Phi}(1)\mathbf{e}_t = \mathbf{\Psi}(1)\boldsymbol{\epsilon}_t$ it follows

$$\mathbf{\Phi}(1)\mathbf{\Lambda}\mathbf{\Phi}(1)' = \mathbf{\Psi}(1)\mathbf{\Sigma}_\epsilon\mathbf{\Psi}(1)'. \tag{16.2}$$

Let $\mathbf{P}$ be a lower triangular matrix of the Cholesky decomposition of $\mathbf{\Psi}(1)\mathbf{\Sigma}_\epsilon\mathbf{\Psi}(1)'$ so that $\mathbf{P}\mathbf{P}' = \mathbf{\Psi}(1)\mathbf{\Sigma}_\epsilon\mathbf{\Psi}(1)'$. Then,

$$\mathbf{\Phi}(1) = \mathbf{P}\mathbf{\Lambda}^{-\frac{1}{2}} \tag{16.3}$$

and

$$\mathbf{\Phi}_0 = \mathbf{\Psi}(1)^{-1}\mathbf{\Phi}(1), \tag{16.4}$$

where $\mathbf{\Lambda} = [diag(\mathbf{P})]^2$. Lastrapes and Selgin (1995) apply this model to study liquidity effects using $\mathbf{y}_t = [r_t, y_t, (m_t - p_t), m_t]'$.

Galí (1999) uses similar long-run restrictions to identify shocks. The main methodological difference from Blanchard and Quah is that Gali uses different variables, log productivity and log hours. Log productivity replaces log GDP. Log hours (or the first difference of log hours) replaces the unemployment rate. The log GDP and unemployment rate used by Blanchard and Quah can lead shocks such as government purchases and permanent labor-supply shocks to be mislabeled as the technological shock. Gali defines correlation of two variables when all shocks but one are shut down as conditional correlation. The estimated conditional correlations of hours and productivity are negative for nontechnology shocks. Hours how a persistent decline in response to a positive technology shock. These findings are hard to reconcile with a RBC model, but are consistent with a model with monopolistic competition and sticky price.

16.1.2 Short-run and Long-Run Restrictions for Structural VAR Models

Galí (1992) uses both short-run and long-run restrictions to identify a structural VAR. He considers an IS-LM model that consists of output (y_t), money supply (m_t), the nominal interest rate (r_t), and the price level (p_t)[1]:

$$\mathbf{B}(L)\mathbf{y}_t = \boldsymbol{\delta} + \mathbf{e}_t \tag{16.5}$$

where $\mathbf{B}(L) = \mathbf{B}_0 - \sum_{i=1}^{p} \mathbf{B}_i L^i$, $\mathbf{B}_0$ has ones on its diagonal, $\mathbf{y}_t = (\Delta y_t, \Delta r_t, r_t - \Delta p_t, \Delta m_t - \Delta p_t)'$, p is the lag order of VAR, L is the lag operator, and $\mathbf{e}_t = (e_t^s, e_t^{ms}, e_t^{md}, e_t^{is})'$ is the vector stochastic process describing supply, money supply, money demand, and spending (IS) disturbances that are assumed to be serially uncorrelated. Let n denote the dimension of $\mathbf{y}_t$, that is, $n = 4$ in this model.

The model (16.5) can be estimated by the reduced form VAR:

$$\mathbf{A}(L)\mathbf{y}_t = \boldsymbol{\delta}_\epsilon + \boldsymbol{\epsilon}_t \tag{16.6}$$

where $\mathbf{A}(L) = \mathbf{I} - \sum_{i=1}^{p} \mathbf{A}_i L^i$, $\mathbf{A}_0 = \mathbf{I}$, and $\boldsymbol{\epsilon}_t$ is the vector of innovations in the elements of $\mathbf{y}_t$. Let $\boldsymbol{\Sigma}_\epsilon$ denote the variance-covariance matrix of $\boldsymbol{\epsilon}_t$. Provided that $\mathbf{B}_0$ is identified, all the structural parameters in (16.5) are computed from the estimates of (16.6) using $\boldsymbol{\delta} = \mathbf{B}_0\boldsymbol{\delta}_\epsilon$ and $\mathbf{B}_i = \mathbf{B}_0\mathbf{A}_i$ for $i = 1, 2, \cdots, p$. Structural shocks are also constructed by $\mathbf{e}_t = \mathbf{B}_0\boldsymbol{\epsilon}_t$.

[1] y_t, m_t, and p_t are in logarithms.

In order to identify $\mathbf{B}_0$, Galí (1992) imposes an orthogonality condition ($\mathcal{R}0$) that the variance-covariance matrix of structural shocks, $\mathbf{\Lambda}$, is diagonal. From $\mathbf{B}_0\mathbf{\Sigma}_\epsilon\mathbf{B}_0' = \mathbf{\Lambda}$ we have $\frac{n(n+1)}{2} = 10$ independent restrictions, and leave $\frac{n(n-1)}{2} = 6$ free parameters in $\mathbf{B}_0$.

A second set of restrictions, building on Blanchard and Quah (1989), specifies that the supply shock has long-run effects on the level of output but the three aggregate demand shocks (e_t^{ms}, e_t^{md}, and e_t^{is}) have no long-run effects on the level of output ($\mathcal{R}1$, $\mathcal{R}2$, and $\mathcal{R}3$). These restrictions identify the supply shock (e_t^s) from the other shocks. These restrictions are denoted by $\mathbf{\Phi}(1)_{1j} = 0$ for $j = 2, 3$, and 4.

A third set of restrictions is that the money supply and the money demand shocks have no contemporaneous effects on output ($\mathcal{R}4$ and $\mathcal{R}5$). These restrictions identify the IS shock from the two types of monetary shocks. Let $\mathbf{\Phi}(L) = \mathbf{B}(L)^{-1}$, in particular, $\mathbf{\Phi}_0 = \mathbf{B}_0^{-1}$. These two restrictions are denoted by $\mathbf{\Phi}_{0,1j} = 0$ for $j = 2$ and 3.

The final restriction identifies the money supply shock from the money demand shock. Galí (1992) assumes that the contemporaneous price does not enter the money supply rule that is denoted by $\mathbf{B}_{0,23} + \mathbf{B}_{0,24} = 0$ ($\mathcal{R}6$).[2]

The estimation of Galí (1992) is dramatic, and is well described by Pagan and Robertson (1995, 1998). From the long-run restrictions ($\mathcal{R}1 \sim \mathcal{R}3$), $\mathbf{\Phi}(1)$ becomes a block lower triangular matrix, where $\mathbf{\Phi}(L) = \mathbf{B}(L)^{-1}$ in (16.5). Inverting $\mathbf{\Phi}(1)$, we also have a block lower triangular matrix $\mathbf{B}(1)$ so that $\mathbf{B}_{12}(1) = \mathbf{B}_{13}(1) = \mathbf{B}_{14}(1) = 0$. We can impose this set of restrictions directly on the coefficients of the structural

[2] Galí (1992) suggests two more alternative assumptions; contemporaneous output does not enter the money supply rule ($\mathcal{R}7$) and contemporaneous homogeneity in money demand ($\mathcal{R}8$). In this section, we focus on ($\mathcal{R}6$).

VAR. For notational convention, let b_{ij} and $b_{s,ij}$ be the (i,j) components of $\mathbf{B}_0$ and $\mathbf{B}_s$, respectively. By imposing these long-run restrictions ($\mathcal{R}1 \sim \mathcal{R}3$), we can reparameterize the first equation of (16.5) as

$$(16.7) \qquad y_{1t} = -b_{12}\Delta^p y_{2t} - b_{13}\Delta^p y_{3t} - b_{14}\Delta^p y_{4t} + \sum_{i=1}^{p} b_{i,11} y_{1,t-i} + \sum_{i=1}^{p-1} b_{i,12}\Delta^{p-i} y_{2t-i} + \sum_{i=1}^{p-1} b_{i,13}\Delta^{p-i} y_{3t-i} + \sum_{i=1}^{p-1} b_{i,14}\Delta^{p-i} y_{4t-i} + e_{1t},$$

where $\Delta^p y_{2t}$ is, for example, $y_{2t} - y_{2,t-p}$, and estimate the coefficients by instrumental variables using y_{it-1} for $\Delta^p y_{it}$ for $i = 2, 3, 4$. Similarly, with the short-run restriction ($\mathcal{R}6$), we can reparameterize the second equation of (16.5) as

$$(16.8) \qquad y_{2t} = -b_{21} y_{1t} - b_{23}(y_{3t} - y_{4t}) + \sum_{i=1}^{p} b_{i,21} y_{1,t-i} + \sum_{i=1}^{p} b_{i,22} y_{2,t-i} + \sum_{i=1}^{p} b_{i,23} y_{3,t-i} + \sum_{i=1}^{p} b_{i,24} y_{4,t-i} + e_{2t},$$

where we use $\hat{\epsilon}_{1t}$, a sample counterpart of the first error in (16.6) from a reduced form VAR, and $\hat{e}_{1t}$, a sample counterpart of the first shock in (16.7) from a structural VAR, for y_{1t} and $y_{3t} - y_{4t}$ as an instrument, respectively. This result follows because ϵ_{1t} is orthogonal to e_{2t} by the short-run restriction ($\mathcal{R}4$) and e_{1t} is orthogonal to e_{2t} by the orthogonality conditions. The third equation is given by

$$(16.9) \qquad y_{3t} = -b_{31} y_{1t} - b_{32} y_{2t} - b_{34} y_{4t} + \sum_{i=1}^{p} b_{i,31} y_{1,t-i} + \sum_{i=1}^{p} b_{i,32} y_{2,t-i} + \sum_{i=1}^{p} b_{i,33} y_{3,t-i} + \sum_{i=1}^{p} b_{i,34} y_{4,t-i} + e_{3t},$$

where $\hat{\epsilon}_{1t}, \hat{e}_{1t}$, and $\hat{e}_{2t}$ are used as the instrumental variables for y_{1t}, y_{2t}, and y_{4t}, respectively. The short-run restriction ($\mathcal{R}5$) ensures that ϵ_{1t} is orthogonal to e_{3t}, while the orthogonality conditions are used for e_{1t} and e_{2t}. Finally, the fourth equation is given by

$$(16.10)\quad y_{4t} = -b_{41}y_{1t} - b_{42}y_{2t} - b_{43}y_{3t} + \sum_{i=1}^{p} b_{i,41}y_{1,t-i} + \sum_{i=1}^{p} b_{i,42}y_{2,t-i} + \sum_{i=1}^{p} b_{i,43}y_{3,t-i} + \sum_{i=1}^{p} b_{i,44}y_{4,t-i} + e_{4t}$$

and estimated by instrumental variables using $\hat{e}_{1t}, \hat{e}_{2t}$, and $\hat{e}_{3t}$ for y_{1t}, y_{2t}, and y_{3t}, respectively from the orthogonality conditions.

The estimation method described above is a two-step instrumental variables method because the reduced form VAR is estimated in the first step and some of the residuals estimated in the first step are used for instrumental variables in the second step.

16.2 Representations for the Cointegrated System

This section introduces four useful representations of a cointegrating system: the *vector moving average representation* and *Phillips' triangular representation.* For example, these representations are useful in developing different methods to impose long-run restrictions.[3] For the illustration below, consider a vector of difference stationary processes $\mathbf{z}_t = (\mathbf{y}_t, \mathbf{x}_t)'$ with a cointegrating vector $\boldsymbol{\beta} = (\mathbf{I}, -\mathbf{c}')'$.

16.2.1 Vector Moving Average Representation

The cointegrating relationship between $\mathbf{y}_t$ and $\mathbf{x}$, and the difference stationarity of $\mathbf{x}_t$ can be written as

$$(16.11)\quad \mathbf{y}_t = \mathbf{c}'\mathbf{x}_t + \mathbf{u}_t$$

$$(16.12)\quad \mathbf{x}_t = \mathbf{x}_{t-1} + \mathbf{v}_t,$$

[3]Details of these representations are discussed in Section 19.1 of Hamilton (1994).

where $\mathbf{u}_t$ and $\mathbf{v}_t$ are stationary with zero mean.

Differencing (16.11) yields

$$\Delta\mathbf{y}_t = \mathbf{c}'\Delta\mathbf{x}_t + \Delta\mathbf{u}_t = \mathbf{c}'\mathbf{v}_t + \mathbf{u}_t - \mathbf{u}_{t-1}. \tag{16.13}$$

Let $\mathbf{e}_{1,t} \equiv \mathbf{c}'\mathbf{v}_t + \mathbf{u}_t$ and $\mathbf{e}_{2,t} \equiv \mathbf{v}_t$. Then, (16.56) can be written as

$$\Delta\mathbf{y}_t = \mathbf{e}_{1,t} - (\mathbf{e}_{1,t-1} - \mathbf{c}'\mathbf{e}_{2,t-1}) = (\mathbf{I} - L)\mathbf{e}_{1,t} + \mathbf{c}'L\mathbf{e}_{2,t}.$$

Stacking this along with (16.12) in a vector system yields the vector moving average representation for $(\Delta\mathbf{y}_t, \Delta\mathbf{x}_t)'$,

$$\begin{bmatrix} \Delta\mathbf{y}_t \\ \Delta\mathbf{x}_t \end{bmatrix} = \mathbf{\Phi}(L) \begin{bmatrix} \mathbf{e}_{1,t} \\ \mathbf{e}_{2,t} \end{bmatrix},$$

where

$$\mathbf{\Phi}(L) \equiv \begin{bmatrix} \mathbf{I} - L & \mathbf{c}'L \\ \mathbf{0} & \mathbf{I} \end{bmatrix}.$$

Note that the polynomial $\mathbf{\Phi}(z)$ has a root at unity, $|\mathbf{\Phi}(1)| = \mathbf{0}$, and hence is non-invertible. This suggests that $\Delta\mathbf{z}_t$ cannot be represented by any finite-order vector autoregression since $[\mathbf{\Phi}(L)]^{-1}\Delta\mathbf{z}_t = \mathbf{e}_t$ does not exist.

Stationarity of $\boldsymbol{\beta}'\mathbf{z}_t$ requires that the vector moving average representation satisfies two necessary conditions. First, the matrix polynomial associated with the moving average must satisfy

$$\boldsymbol{\beta}'\mathbf{\Phi}(1) = \mathbf{0}.$$

Further, if some of the series in $\mathbf{z}_t$ exhibit nonzero drift and thus include the deterministic trend component $\boldsymbol{\mu}_z t$,

$$\mathbf{z}_t = \boldsymbol{\mu}_z t + \mathbf{z}_t^0,$$

where $\boldsymbol{\mu}_z \neq \mathbf{0}$, and $\mathbf{z}_t^0$ is difference stationary without drift, then the stationarity requires that the deterministic cointegration restriction holds (Engle and Yoo, 1987;

Ogaki and Park, 1997). That is, the cointegrating vector must eliminate the deterministic trend from the system:

$$\boldsymbol{\beta}'\boldsymbol{\mu}_z = \mathbf{0}.$$

Otherwise, the linear combination $\boldsymbol{\beta}'\mathbf{z}_t$ will grow deterministically at the rate $\boldsymbol{\beta}'\boldsymbol{\mu}_z$.

16.2.2 Phillips' Triangular Representation

Phillips's (1991) triangular representation takes the form:

$$\mathbf{y}_t - \mathbf{c}'\mathbf{x}_t = \mathbf{u}_t, \tag{16.14}$$

$$\Delta\mathbf{x}_t = \mathbf{v}_t. \tag{16.15}$$

To derive this, suppose an $n\times 1$ vector $\mathbf{z}_t = (\mathbf{y}_t, \mathbf{x}_t)'$ is characterized by h cointegrating relations. The matrix of h cointegrating vectors can be written as

$$\boldsymbol{\beta}' = \begin{bmatrix} \mathbf{b}_1' \\ \mathbf{b}_2' \\ \vdots \\ \mathbf{b}_h' \end{bmatrix} = \begin{bmatrix} 1 & b_{12} & b_{13} & \cdots & b_{1n} \\ b_{21} & b_{22} & b_{23} & \cdots & b_{2n} \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ b_{h1} & b_{h2} & b_{h3} & \cdots & b_{hn} \end{bmatrix},$$

where the (1,1)-th element has been normalized to unity. After appropriate row operations, it can be transformed as

$$\boldsymbol{\beta}' = \begin{bmatrix} 1 & 0 & \cdots & 0 & b^*_{1,h+1} & b^*_{1,h+2} & \cdots & b^*_{1,n} \\ 0 & 1 & \cdots & 0 & b^*_{2,h+1} & b^*_{2,h+2} & \cdots & b^*_{2,n} \\ \vdots & \vdots & \cdots & \vdots & \vdots & \vdots & \cdots & \vdots \\ 0 & 0 & \cdots & 1 & b^*_{h,h+1} & b^*_{h,h+2} & \cdots & b^*_{h,n} \end{bmatrix} = \begin{bmatrix} \mathbf{I}_h & -\mathbf{c}' \end{bmatrix}.$$

Therefore, with $\mathbf{z}_t$ correspondingly partitioned into an $h\times 1$ vector $\mathbf{y}_t$ and a $(n-h)\times 1$ vector $\mathbf{x}_t$,

$$\boldsymbol{\beta}'\mathbf{z}_t = \begin{bmatrix} \mathbf{I}_h & -\mathbf{c}' \end{bmatrix} \begin{bmatrix} \mathbf{y}_t \\ \mathbf{x}_t \end{bmatrix} = \mathbf{y}_t - \mathbf{c}'\mathbf{x}_t$$

is stationary in equation (16.57). Equation (16.58) comes from the assumption that $\mathbf{z}_t$ is difference stationary. Thus, in Phillips' triangular representation, variables on the left hand side are all stationary, and are expressed in the form of the moving average.

The triangular representation has been widely used for estimating cointegrating vectors. One of the reasons is that when presented in this way, the model's (unknown) coefficients appear only in equation (16.57). Therefore, we can estimate the cointegrating relationship using standard estimation methods for a system of simultaneous equations.

As an example of Phillips' representation, consider the 4-variable system of Shapiro and Watson (1988). The model consists of four variables: labor input h_t, output y_t, the inflation rate π_t, and the long-run real interest rate $i_t - \pi_t$. In the short-run, these variables deviate from their long-run steady state values due to four types of serially uncorrelated shocks: labor supply shocks v_t, technological shocks e_t, and two aggregate demand shocks ν_t^1 and ν_t^2. Labor supply shocks and technology shocks are uncorrelated with each other and with the aggregate demand shocks. In this model, all shocks are assumed to have only short-term effects on the real interest rate. That is, the nominal interest rate and the inflation rate are cointegrated so the real interest rate is stationary. Let

$$\mathbf{z}_t = [\ i_t \quad \pi_t \quad h_t \quad y_t\ ]',$$

with a cointegrating vector

$$\boldsymbol{\beta}' = [\ 1 \quad -1 \quad 0 \quad 0\ ].$$

We can partition $\mathbf{z}_t$ into $z_{1,t} = i_t$, and $\mathbf{z}_{2,t} = (\ \pi_t \quad h_t \quad y_t\ )'$. With the model's long-run restrictions, Phillips' triangular representation for this cointegrating system is

$$\begin{aligned}
i_t - \pi_t &= c_1 + \mathbf{\Phi}_i(L)[\ v_t \quad e_t \quad \nu_t^1 \quad \nu_t^2\ ]', \\
\Delta\pi_t &= c_2 + \mathbf{\Phi}_\pi(L)[\ v_t \quad e_t \quad \nu_t^1 \quad \nu_t^2\ ]', \\
\Delta h_t &= c_3 + \mathbf{\Sigma}_h(L)v_t + (1-L)\mathbf{\Phi}_h(L)[\ v_t \quad e_t \quad \nu_t^1 \quad \nu_t^2\ ]', \\
\Delta y_t &= c_4 + \mathbf{\Sigma}_h(L)v_t + \alpha^{-1}\mathbf{\Sigma}_e(L)e_t + (1-L)\mathbf{\Phi}_y(L)[\ v_t \quad e_t \quad \nu_t^1 \quad \nu_t^2\ ]',
\end{aligned}$$

where c_i for $i = 1, \cdots, 4$, are constant, and the lag polynomials $\mathbf{\Sigma}_h(L)$ and $\mathbf{\Sigma}_\varepsilon(L)$ are assumed to have absolutely summable coefficients and roots outside the unit circle.

16.2.3 Vector Error Correction Model Representation

Vector autoregressive models originating with Sims (1980) have the following reduced form:

$$\mathbf{A}(L)\mathbf{y}_t = \boldsymbol{\delta}_\epsilon + \boldsymbol{\epsilon}_t, \tag{16.16}$$

where $\mathbf{A}(L) = \mathbf{I}_n - \sum_{i=1}^{p} \mathbf{A}_i L^i$, $\mathbf{A}(0) = \mathbf{I}_n$, and $\boldsymbol{\epsilon}_t$ is *white noise* with mean zero and variance $\mathbf{\Sigma}_\epsilon$. From the reduced form of the VAR model, $\mathbf{A}(L)$ can be re-parameterized as $\mathbf{A}(1)L + \mathbf{A}^*(L)(1-L)$, where $\mathbf{A}(1)$ has a reduced rank, $r < n$. Engle and Granger (1987) showed that there exists an error correction representation:

$$\mathbf{A}^*(L)\Delta\mathbf{y}_t = \boldsymbol{\delta}_\epsilon - \mathbf{A}(1)\mathbf{y}_{t-1} + \boldsymbol{\epsilon}_t, \tag{16.17}$$

where $\mathbf{A}^*(L) = \mathbf{I}_n - \sum_{i=1}^{p-1} \mathbf{A}_i^* L^i$, and $\mathbf{A}_i^* = -\sum_{j=i+1}^{p} \mathbf{A}_j$. Since $\mathbf{y}_t$ is assumed to be cointegrated $I(1)$, $\Delta\mathbf{y}_t$ is $I(0)$, and $-\mathbf{A}(1)$ can be decomposed as $\boldsymbol{\alpha\beta}'$, where $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ are $n \times r$ matrices with full column rank, r.

Monte Carlo experiments of Qureshi (2008) show that for OLS estimates of level VAR very often exhibit explosive autoregressive roots for typical macro data. In contrast, the frequency of encountering explosive roots in OLS estimates of VECM is much fewer. Because there is a general consensus among macroeconomists that the absolute value of autroregessive roots is at most one, this is an important advantage for VECM over level VAR.

16.2.4 Common Trend Representation

Another representation of a cointegrated VAR system is Stock and Watson (1988b) common trend representation, which is a generalization of Beverage-Nelson decomposition. Since Δy_t is stationary, we have

$$(1 - L)\mathbf{y_t} = \mathbf{\Phi}(L)\boldsymbol{\epsilon}_t. \tag{16.18}$$

Then

$$\begin{aligned} \mathbf{y_t} &= \frac{\mathbf{\Phi}(L)}{1-L} \\ &= \frac{\mathbf{\Phi}(1)}{1-L}\boldsymbol{\epsilon}_t + \frac{\mathbf{\Phi}(L) - \mathbf{\Phi}(1)}{1-L}\boldsymbol{\epsilon}_t \\ &= A\begin{bmatrix} z_{1,t} \\ \vdots \\ z_{n-r,t} \end{bmatrix} + \boldsymbol{B}(L)\boldsymbol{\epsilon}_t \end{aligned} \tag{16.19}$$

where $z_{i,t}$ is a random walk and is called a stochastic trend. In a n-variable system, there exist r cointegration relationship if and only if there exist $(n - r)$ common stochastic trend.

Example 16.1 If we have income and consumption, y_t and c_t, such that

$$y_t = z_t + e_t^y \tag{16.20}$$

$$c_t = z_t + e_t^c \tag{16.21}$$

where z_t is a random walk, and e_t^y and e_t^c are transitory income and consumption shock, respectively. Then,

$$\begin{pmatrix} y_t \\ c_t \end{pmatrix} = z_t + \begin{pmatrix} e_t^y \\ e_t^c \end{pmatrix}. \tag{16.22}$$

where z_t is a common stochastic trend. In this case, there is one cointegrating relationship so that $y_t - c_t = e_t^y - e_t^c$ is stationary. ∎

16.3 Long-Run Restrictions on Phillips' Triangular Representation

Long-run restrictions can be imposed on Phillips' Triangular representation. As an illustration, consider the model of Shapiro and Watson (1988). In this model, $\mathbf{y}_t = (\Delta h_t, \Delta y_t, \Delta \pi_t, i_t - \pi_t)'$, where h_t denotes labor supply, y_t output, π_t inflation, and i_t the nominal interest rate. Since h_t, y_t, and π_t are assumed to be $I(1)$, Δh_t, Δy_t, and $\Delta \pi_t$ are stationary $I(0)$. There are three sources of disturbances: labor supply v_t, technology e_t, and aggregate demand disturbances ν_t^1 and ν_t^2, and thus $\mathbf{e}_t = (v_t, e_t, \nu_t^1, \nu_t^2)'$. The first two disturbances may be referred as supply shocks, and are assumed to be orthogonal and serially uncorrelated, and uncorrelated with the demand shocks. Since $\mathbf{y}_t$ has been assumed to be stationary, none of the shocks has a long-run effect on Δh_t, Δy_t, $\Delta \pi_t$, or $i_t - \pi_t$.

Shapiro and Watson (1988) make two identifying restrictions: first, the aggregate demand shocks have no permanent effect on the level of output; and second, the long-run level of labor supply is exogenous. To impose these restrictions, consider, for example, the long-run effect of ν_t^1 on y_t. In their setup, ϕ_{23k} is the effect of ν_t^1 on Δy_t after k periods, and therefore $\sum_{k=1}^{l} \phi_{23k}$ is the effect of ν_t^1 on y_t itself after l periods. For ν_t^1 to have no effect on y_t in the long run, then we must have that $\sum_{k=0}^{\infty} \phi_{23k} = 0$.

Thus, the two assumptions impose restrictions that the long-run multipliers from ν_t^1 and ν_t^2 to h_t and y_t, and from e_t to h_t are zero. The resulting matrix of long-run multipliers, $\Phi(1)$, is block lower triangular:

$$\Phi(1) = \begin{bmatrix} \phi_{11} & 0 & 0 & 0 \\ \phi_{21} & \phi_{22} & 0 & 0 \\ \phi_{31} & \phi_{32} & \phi_{33} & \phi_{34} \\ \phi_{41} & \phi_{42} & \phi_{43} & \phi_{44} \end{bmatrix}.$$

Because there are no restrictions on ϕ_{34}, this identification scheme cannot be used to disentangle the two aggregate demand shocks ν_t^1 and ν_t^2, and only their joint impact can be estimated.

In order to estimate $\mathbf{e}_t$ and $\Phi(L)$ using the observed data, Shapiro and Watson (1988) follow Blanchard and Quah (1989), and use the block lower triangular structure of $\Phi(1)$ and the assumption that the shocks are serially and mutually uncorrelated. The Wold representation $\mathbf{y}_t = \delta + \Psi(L)\epsilon_t$ can be obtained by first estimating and then inverting the VAR representation of $\mathbf{y}_t$ in the usual way.

The equation for Δh_t can be written as

$$\Delta h_t = \sum_{j=1}^{p} \beta_{hh,j}\Delta h_{t-j} + \sum_{j=0}^{p} \beta_{hy,j}\Delta y_{t-j} + \sum_{j=0}^{p} \beta_{h\pi,j}\Delta \pi_{t-j} + \sum_{j=0}^{p} \beta_{hi,j}(i_{t-j} - \pi_{t-j}) + v_t.$$

Because the long-run multipliers from e_t, ν_t^1, and ν_t^2 to h_t are zero, $\sum_{j=0}^{p} \beta_{hn,j} = 0$ for $n = y, \pi, i$. Imposing these constrains yields second differences. For example, consider the long-run restriction of e_t on h_t:

$$\begin{aligned} \sum_{j=0}^{p} \beta_{hy,j}\Delta y_{t-j} &= \beta_{hy,0}\Delta y_t + \cdots + \beta_{hy,p-1}\Delta y_{t-(p-1)} + \beta_{hy,p}\Delta y_{t-p} \\ &= \beta_{hy,0}(\Delta y_t - \Delta y_{t-1}) + (\beta_{hy,0} + \beta_{hy,1})(\Delta y_{t-1} - \Delta y_{t-2}) + \cdots \\ &\quad +(\beta_{hy,0} + \beta_{hy,1} + \cdots + \beta_{hy,p-1})(\Delta y_{t-(p-1)} - \Delta y_{t-p}) \\ &\quad +(\beta_{hy,0} + \beta_{hy,1} + \cdots + \beta_{hy,p-1} + \beta_{hy,p})(\Delta y_{t-p}) \end{aligned}$$

The long-run restriction requires that $\beta_{hy,0} + \beta_{hy,1} + \cdots + \beta_{hy,p-1} + \beta_{hy,p} = 0$, and hence the coefficient on Δy_{t-p} is zero. Thus we have

$$\begin{aligned}\sum_{j=0}^{p} \beta_{hy,j}\Delta y_{t-j} &= \beta_{hy,0}\Delta^2 y_t + (\beta_{hy,0} + \beta_{hy,1})\Delta^2 y_{t-1} + \cdots + (\beta_{hy,0} + \beta_{hy,1} + \cdots + \beta_{hy,p-1})\Delta^2 y_{t-(p-1)} \\ &= \gamma_{hy,0}\Delta^2 y_t + \gamma_{hy,1}\Delta^2 y_{t-1} + \cdots + \gamma_{hy,p-1}\Delta^2 y_{t-(p-1)} \\ &= \sum_{j=0}^{p-1} \gamma_{hy,s}\Delta^2 y_{t-j}.\end{aligned}$$

The same operations can be done for $\sum_{j=0}^{p} \beta_{h\pi,j}$ and $\sum_{j=0}^{p} \beta_{hi,j}$ as well. The resulting equation to be estimated is

$$\Delta h_t = \sum_{j=1}^{p} \beta_{hh,j}\Delta h_{t-j} + \sum_{j=0}^{p-1} \gamma_{hy,j}\Delta^2 y_{t-j} + \sum_{j=0}^{p-1} \gamma_{h,\pi}\Delta^2 \pi_{t-j} + \sum_{j=0}^{p-1} \gamma_{hi,j}(\Delta i_{t-j} - \Delta\pi_{t-j}) + v_t.$$

This equation cannot be consistently estimated by OLS because it includes contemporaneous values of some of the regressors which are correlated with v_t. Therefore, the IV estimation is used with $\{\Delta h_{t-s}, \Delta y_{t-s}, \Delta\pi_{t-s}, i_{t-s} - \pi_{t-s}\}_{s=1}^{p}$ as instruments.

Similarly, the equation for Δy_t is

$$\Delta y_t = \sum_{j=1}^{p} \beta_{yh,j}\Delta h_{t-j} + \sum_{j=1}^{p} \beta_{yy,j}\Delta y_{t-j} + \sum_{j=0}^{p-1} \Delta^2 \pi_{t-j} + \sum_{j=0}^{p-1} \gamma_{yi,j}(\Delta i_{t-j} - \Delta\pi_{t-j}) + \beta_{yv}v_t + e_t.$$

Note that the contemporaneous value of Δh_t do not enter this equation since v_t enters directly. Again, the correlations between e_t and contemporaneous values of some of the regressors require that it is estimated by the IV estimation using the same set of instruments plus $\{v_{t-s}\}_{s=1}^{p}$ as instruments.

The equations estimated for $\Delta\pi_t$ and $\pi_t - i_t$ are reduced forms. They are

$$\Delta\pi_t = \sum_{j=1}^{p} \beta_{\pi h,j}\Delta h_{t-j} + \sum_{j=0}^{p} \beta_{\pi y,j}\Delta y_{t-j} + \sum_{j=1}^{p} \beta_{\pi\pi,j}\Delta\pi_{t-j} + \sum_{j=1}^{p} \beta_{\pi i,j}(i_{t-j} - \pi_{t-j}) + \beta_{\pi v}v_t + \beta_{\pi e}e_t + a_t^1,$$

and

$$i_t - \pi_t = \sum_{j=1}^{p} \beta_{ih,j}\Delta h_{t-j} + \sum_{j=0}^{p} \beta_{iy,j}\Delta y_{t-j} + \sum_{j=1}^{p} \beta_{i\pi,j}\Delta\pi_{t-j} + \sum_{j=1}^{p} \beta_{ii,j})i_{t-j} - \pi_{t-j}) + \beta_{iv}v_t + \beta_{ie}e_t + a_t^2.$$

The error terms a_t^1 and a_t^2 are linear combinations of the structural aggregate shocks ν_t^1 and ν_t^2. Since these disturbances are uncorrelated with the regressions, these two equations can be estimated by OLS.

16.3.1 Long-run Restrictions and VECM

An alternative method to impose long-run restrictions is to use VECM. As $\Delta \mathbf{y}_t$ is assumed to be stationary, it has a unique Wold representation:

$$\Delta \mathbf{y}_t = \boldsymbol{\mu} + \boldsymbol{\Psi}(L)\boldsymbol{\epsilon}_t, \tag{16.23}$$

where $\boldsymbol{\mu} = \boldsymbol{\Psi}(1)\boldsymbol{\delta}_\epsilon$ and $\boldsymbol{\Psi}(L) = \mathbf{I}_n + \sum_{i=1}^{\infty} \boldsymbol{\Psi}_i L^i$. The above, which is in reduced form, can be represented in structural form as:

$$\begin{aligned} \Delta \mathbf{y}_t &= \boldsymbol{\mu} + \boldsymbol{\Phi}(L)\mathbf{e}_t \\ \boldsymbol{\Phi}(L) &= \boldsymbol{\Psi}(L)\boldsymbol{\Phi}_0 \\ \mathbf{e}_t &= \boldsymbol{\Phi}_0^{-1}\boldsymbol{\epsilon}_t, \end{aligned} \tag{16.24}$$

where $\boldsymbol{\Phi}(L) = \boldsymbol{\Phi}_0 + \sum_{i=1}^{\infty} \boldsymbol{\Phi}_i L^i$, and $\mathbf{e}_t$ is a vector of structural innovations with mean zero and variance $\boldsymbol{\Lambda}$.

Long-run restrictions are imposed on the structural form, as in Blanchard and Quah (1989). Stock and Watson (1988a) developed a common trend representation that was shown to be equivalent to a VECM representation. When cointegrated variables have a reduced rank, r, there exist $k = n - r$ common trends. These common trends can be considered to be generated by permanent shocks, so that $\mathbf{e}_t$ can be decomposed into $(\mathbf{e}_t^{k\prime}, \mathbf{e}_t^{r\prime})'$, in which $\mathbf{e}_t^k$ is a k-dimensional vector of permanent shocks and $\mathbf{e}_t^r$ is an r-dimensional vector of transitory shocks. As developed in King, Plosser, Stock, and Watson (1989, 1991, KPSW for short), this decomposition ensures that

$$\mathbf{\Phi}(1) = \left[\begin{array}{cc} \mathbf{A} & \mathbf{0} \end{array} \right], \tag{16.25}$$

where $\mathbf{A}$ is an $n \times k$ matrix and $\mathbf{0}$ is an $n \times r$ matrix with zeros, representing long-run effects of permanent shocks and transitory shocks, respectively. In order to identify permanent shocks, in general, causal chains, in the sense of Sims (1980), are imposed on permanent shocks:

$$\mathbf{A} = \hat{\mathbf{A}}\mathbf{\Pi}, \tag{16.26}$$

where $\hat{\mathbf{A}}$ is an $n \times k$ matrix, and $\mathbf{\Pi}$ is a $k \times k$ lower triangular matrix with ones in the diagonal. As Jang (2001a) shows, $\hat{\mathbf{A}}$ is constructed using the cointegrating vectors:

$$\hat{\mathbf{A}} = \hat{\boldsymbol{\beta}}_{\perp}. \tag{16.27}$$

See Appendix 16.A for detail.

16.3.2 Identification of Permanent Shocks

The main interest lies in the identification of structural permanent shocks, not in structural transitory shocks.[4] Following KPSW, we decompose $\mathbf{\Phi}_0$ and $\mathbf{\Phi}_0^{-1}$ as:

$$\mathbf{\Phi}_0 = \left[\begin{array}{cc} \mathbf{H} & \mathbf{J} \end{array} \right], \quad \mathbf{\Phi}_0^{-1} = \left[\begin{array}{c} \mathbf{G} \\ \mathbf{E} \end{array} \right] \tag{16.28}$$

where $\mathbf{H}, \mathbf{J}, \mathbf{G}$ and $\mathbf{E}$ are $n \times k$, $n \times r$, $k \times n$, and $r \times n$ matrices, respectively. Note that the permanent shocks are identified once $\mathbf{H}$ (or $\mathbf{G}$) is identified, and that these two matrices have a one-to-one relation, $\mathbf{G} = \mathbf{\Lambda}^k \mathbf{H}' \mathbf{\Sigma}_\epsilon^{-1}$, where $\mathbf{\Lambda}^k$ is the variance-covariance matrix of permanent shocks, $\mathbf{e}_t^k$.[5] Therefore, the above decomposition of $\mathbf{\Phi}_0$ does not generate additional free parameters.

[4]Fisher, Fackler, and Orden (1995) consider the identification of transitory shocks imposing causal chains on transitory shocks.

[5]One can easily derive this relation from the relation $\mathbf{\Phi}_0^{-1}\mathbf{\Sigma}_\epsilon = \mathbf{\Lambda}\mathbf{\Phi}_0'$.

The identifying scheme below basically follows that of KPSW, but enables one to generalize their model as described below. See Jang (2001a) for details. Following KPSW, let $\mathbf{D} = (\hat{\boldsymbol{\beta}}_{\perp}'\hat{\boldsymbol{\beta}}_{\perp})^{-1}\hat{\boldsymbol{\beta}}_{\perp}'\boldsymbol{\Psi}(1)$ and $\mathbf{P}$ be a lower triangular matrix chosen from the Cholesky decomposition of $\mathbf{D}\boldsymbol{\Sigma}_{\epsilon}\mathbf{D}'$. Then $\boldsymbol{\Pi}$ and $\boldsymbol{\Lambda}^k$ are uniquely determined by

(16.29) $$\boldsymbol{\Pi} = \mathbf{P}(\boldsymbol{\Lambda}^k)^{-\frac{1}{2}},$$

where $\boldsymbol{\Lambda}^k = [diag(\mathbf{P})]^2$, and $\mathbf{H}$ and $\mathbf{G}$ are identified by

(16.30) $$\mathbf{H} = \begin{bmatrix} \mathbf{D} \\ \boldsymbol{\alpha}'\boldsymbol{\Sigma}_{\epsilon}^{-1} \end{bmatrix}^{-1} \begin{bmatrix} \boldsymbol{\Pi} \\ \mathbf{0} \end{bmatrix}$$

and

(16.31) $$\mathbf{G} = \boldsymbol{\Lambda}^k\mathbf{H}'\boldsymbol{\Sigma}_{\epsilon}^{-1}.$$

Accordingly, the permanent shocks and the short run dynamics are identified by

(16.32) $$\mathbf{e}_t^k = \mathbf{G}\epsilon_t$$

and

(16.33) $$\boldsymbol{\Phi}(L)^k = \boldsymbol{\Psi}(L)\mathbf{H},$$

where $\boldsymbol{\Phi}(L)^k$ denotes the first k columns of $\boldsymbol{\Phi}(L)$.

The specific solutions for $\mathbf{H}$ and $\mathbf{G}$ in the form of matrices enable one to generalize the model. Jang (2001b) considered a structural VECM in which structural shocks are partially identified using long-run restrictions and are fully identified by means of additional short-run restrictions (See Jang, 2001b, for the method of identification in structural VECMs with short-run and long-run restrictions). Jang and Ogaki (2001) consider a special case, where impulse response analysis is used to examine the effects of only one permanent shock, and the recursive assumption on the permanent shocks in (16.26) can be relaxed, which implies $\mathbf{\Pi}$ is lower block triangular. Note that we can compute the impulse responses to the k_{th} shock as long as the k_{th} column of $\mathbf{H}$, $\mathbf{H}_k$, is identified. Note also that the third column of $\mathbf{\Pi}$ does not contain any unknown parameters. Analogous to (16.30), $\mathbf{H}_k$ is identified by

$$\mathbf{H}_k = \begin{bmatrix} \mathbf{D} \\ \boldsymbol{\alpha}'\mathbf{\Sigma}_\epsilon^{-1} \end{bmatrix}^{-1} \mathbf{S}_k \tag{16.34}$$

where $\mathbf{S}_k$ is an n-dimensional selection vector with one at the k_{th} row and zeros at other rows. Similarly, $\mathbf{G}_k$ is identified by:

$$\mathbf{G}_k = \mathbf{\Lambda}_{k,k}^k \mathbf{H}_k' \mathbf{\Sigma}_\epsilon^{-1} \tag{16.35}$$

and it follows from the identity relation of $\mathbf{GH} = \mathbf{I}_k$ that

$$\mathbf{\Lambda}_{k,k}^k = (\mathbf{H}_k' \mathbf{\Sigma}_\epsilon^{-1} \mathbf{H}_k)^{-1}, \tag{16.36}$$

where $\mathbf{\Lambda}_{k,k}^k$ is the variance of the k_{th} permanent shock. Thus, the k_{th} permanent shock is identified by

$$e_{t,k}^k = \mathbf{G}_k \boldsymbol{\epsilon}_t. \tag{16.37}$$

16.3.3 Impulse Response Functions

Impulse response analysis has been widely used in the applied VAR literature. It is, however, not straightforward to compute the impulse response from VECMs. The reduced-form VECM is usually converted to a *levels* VAR model for impulse response analysis.[6] Noting that the presence of unit roots prevents the inversion of a *levels*

[6] Mellander, Vredin, and Warne (1992) provide an algorithm to compute impulse response without converting VECM to *levels* VAR following the scheme in Campbell and Shiller (1988) and Warne (1991).

VAR model to a moving average (MA) representation, Lütkepohl and Reimers (1992) suggested the following algorithm to get impulse responses recursively in a cointegrated system. First, estimate the reduced-form VECM in (16.17), then convert the VECM to a *levels* VAR representation in (16.16) using the following relations:[7]

$$\mathbf{A}_i = \begin{cases} \mathbf{I}_n - \mathbf{A}(1) + \mathbf{A}_1^* & i = 1 \\ \mathbf{A}_i^* - \mathbf{A}_{i-1}^* \quad \quad for & 2 \leq i \leq p-1 \\ -\mathbf{A}_{p-1}^* & i = p. \end{cases} \tag{16.38}$$

Though a Wold representation does not exist in the presence of unit roots, Lütkepohl and Reimers (1992) showed that impulse responses can be recursively computed by

$$\mathbf{\Psi}_m = \sum_{l=1}^{p} \mathbf{\Psi}_{m-l}\mathbf{A}_l, \qquad m = 1, 2, 3, \cdots \tag{16.39}$$

$$\mathbf{\Phi}_m = \mathbf{\Psi}_m \mathbf{\Phi}_0, \tag{16.40}$$

where $\mathbf{\Psi}_0 = \mathbf{I}_n$, $\mathbf{\Phi}_m = (\phi_{m,ij})$, and $\phi_{m,ij}$ is an m-step response of the i_{th} variable to the j_{th} innovation.[8] In particular, the impulse response function of permanent shocks in this paper is calculated by[9]

$$\mathbf{\Phi}_m^k = \mathbf{\Psi}_m \mathbf{H}, \qquad m = 1, 2, \cdots . \tag{16.41}$$

As a special case, discussed in Section 16.3.2, the impulse response function of the k_{th} permanent shock is uniquely calculated from

$$\mathbf{\Phi}_{m,k}^k = \mathbf{\Psi}_m \mathbf{H}_k, \qquad m = 1, 2, \cdots \tag{16.42}$$

where $\mathbf{\Phi}_{m,k}^k$ is equivalent to the k_{th} column of $\mathbf{\Phi}_m^k$ in (16.41).

[7]We assume that $n > p$ without any loss of generality.

[8]This algorithm can be simplified by rewriting VAR in (16.16) as a companion VAR(1) form. Then, $\mathbf{\Psi}_m$ is the first n row and n column submatrix of $\mathbf{A}_c^m$, in which $\mathbf{A}_c$ is a companion form coefficient matrix.

[9]One may calculate the impulse response to a one standard deviation permanent shock by $\mathbf{\Psi}_m \mathbf{H}(\mathbf{\Lambda}^k)^{\frac{1}{2}}$.

16.3.4 Forecast-Error Variance Decomposition

Denoting the h-step forecast error by

$$(16.43)\qquad \begin{aligned}\mathbf{y}_{t+h} - E_t\mathbf{y}_{t+h} &= \sum_{i=0}^{\infty} \mathbf{\Psi}_i(\boldsymbol{\epsilon}_{t+h-i} - E_t\boldsymbol{\epsilon}_{t+h-i}) \\ &= \sum_{i=0}^{h-1} \mathbf{\Psi}_i\boldsymbol{\epsilon}_{t+h-i},\end{aligned}$$

the forecast error variance is computed by the diagonal components of

$$(16.44)\qquad E(\mathbf{y}_{t+h} - E_t\mathbf{y}_{t+h})^2 = \sum_{i=0}^{h-1} \mathbf{\Psi}_i\mathbf{\Sigma}_\epsilon\mathbf{\Psi}_i'.$$

In particular, the forecast error variance of the l_{th} variable, $y_{l,t+h}$, is computed by

$$(16.45)\qquad \sum_{i=0}^{h-1} \mathbf{\Psi}_{i,l\cdot}\mathbf{\Sigma}_\epsilon\mathbf{\Psi}_{i,l\cdot}'.$$

where $\mathbf{\Psi}_{i,l\cdot}$ denotes the l_{th} row of $\mathbf{\Psi}_i$.

To isolate the fraction of the forecast error variance attributed to permanent shocks, it is convenient and necessary to decompose the contribution of permanent shocks and transitory shocks as follows:

$$(16.46)\qquad \begin{aligned}\mathbf{y}_{t+h} - E_t\mathbf{y}_{t+h} &= \sum_{i=0}^{\infty} \mathbf{\Psi}_i\mathbf{\Phi}_0(\mathbf{e}_{t+h-i} - E_t\mathbf{e}_{t+h-i}) \\ &= \sum_{i=0}^{h-1} \mathbf{\Psi}_i \begin{bmatrix} \mathbf{H} & \mathbf{J} \end{bmatrix} \begin{bmatrix} \mathbf{e}_{t+h-i}^k \\ \mathbf{e}_{t+h-i}^r \end{bmatrix},\end{aligned}$$

where $\mathbf{\Psi}_i$ is defined in (16.39). Since $\mathbf{e}_t$ is serially uncorrelated,

$$(16.47)\qquad \begin{aligned}E(\mathbf{y}_{t+h} - E_t\mathbf{y}_{t+h})^2 &= \sum_{i=0}^{h-1} \mathbf{\Psi}_i \begin{bmatrix} \mathbf{H} & \mathbf{J} \end{bmatrix} \begin{bmatrix} \mathbf{\Lambda}^k & \mathbf{0} \\ \mathbf{0} & \mathbf{\Lambda}^r \end{bmatrix} \begin{bmatrix} \mathbf{H}' \\ \mathbf{J}' \end{bmatrix} \mathbf{\Psi}_i' \\ &= \sum_{i=0}^{h-1} \mathbf{\Psi}_i(\mathbf{H}\mathbf{\Lambda}^k\mathbf{H}' + \mathbf{J}\mathbf{\Lambda}^r\mathbf{J}')\mathbf{\Psi}_i'.\end{aligned}$$

Therefore, the contribution of permanent shocks to forecast error variance of the h-step forecast is estimated by the diagonal components of

$$\sum_{i=0}^{h-1} \mathbf{\Phi}_i^k \mathbf{\Lambda}^k \mathbf{\Phi}_i^{k\prime}. \tag{16.48}$$

In particular, the contribution of the m_{th} permanent shock, e_m^k, to the forecast error variance of the l_{th} variable, $y_{l,t+h}$, is[10]

$$\sum_{i=0}^{h-1} (\mathbf{\Phi}_{i,lm}^k)^2 \mathbf{\Lambda}_{m,m}^k, \tag{16.49}$$

where $\mathbf{\Lambda}_{m,m}^k$ is the variance of the m_{th} permanent shock.

Finally, dividing (16.49) by (16.45) yields the fraction of the h-step forecast error variance of the l_{th} variable attributed to the m_{th} structural shock.

Section 16.3.2 discusses the special case of the contribution of the k_{th} permanent shock, e_k^k, to the forecast error variance of the l_{th} variable, $y_{l,t+h}$, which is computed by

$$\sum_{i=0}^{h-1} (\mathbf{\Phi}_{i,lk}^k)^2 \mathbf{\Lambda}_{k,k}^k \tag{16.50}$$

where $\mathbf{\Lambda}_{k,k}^k$ is the variance of the k_{th} permanent shock. Dividing (16.50) by (16.45) gives the portion of the contribution of the k_{th} structural shock to the h-step forecast error variance of the l_{th} variable.

16.3.5 Summary

In summary, the estimation and identification of VECM with long-run restrictions are executed by the following procedure:

1. Select the lag length of VECM using some criteria such as AIC and BIC.

[10] By the virtue of the assumption that permanent shocks are uncorrelated mutually, we can separate the contribution of each permanent shock.

2. Estimate cointegrating vectors and determine the rank of cointegrating vectors in (16.17).

3. Convert VECM to levels VAR using (16.38).

4. Impose long-run restrictions implied by economic theory[11], and identify structural parameters using (16.30) and (16.31).

5. Compute impulse responses to a structural shock using (16.41).

6. Compute forecast-error variance decompositions using (16.45) and (16.49).

7. Compute confidence intervals of impulse responses and standard errors of forecast-error variance decompositions using Monte Carlo integration as described in Appendix 16.B.

16.4 Structural Vector Error Correction Models

In this section, we introduce ECM. Let $\mathbf{y}_t$ be an n-dimensional vector of first difference stationary and stationary random variables. Let $\boldsymbol{\ell}_i = (0, ...0, 1, 0, ...0)'$ with 1 on the i_{th} element. If the i_{th} element of $\mathbf{y}_t$ is stationary, then $\boldsymbol{\ell}_i \mathbf{y}_t$ is stationary. When a time series includes stationary variables, we extend the definition of cointegration, and say that $\mathbf{y}_t$ is cointegrated with $\boldsymbol{\ell}_i$ as a cointegrating vector. Suppose that $\mathbf{y}_t$ has a VAR representation

$$\mathbf{y}_t = \boldsymbol{\delta}_\epsilon + \mathbf{A}_1 \mathbf{y}_{t-1} + \mathbf{A}_2 \mathbf{y}_{t-2} + \cdots + \mathbf{A}_p \mathbf{y}_{t-p} + \boldsymbol{\epsilon}_t. \tag{16.51}$$

[11] For example, one may adopt a long-run restriction that a monetary shock does not affect the level of real output.

where $\boldsymbol{\delta}_\epsilon$ is an $n \times 1$ vector. Just as in Said-Dickey's reparameterization for the univariate case, it is convenient to reparameterize Equation (16.51) as

$$\Delta \mathbf{y}_t = \boldsymbol{\delta}_\epsilon - \mathbf{A}(1)\mathbf{y}_{t-1} + \mathbf{A}_1^* \Delta \mathbf{y}_{t-1} + \cdots + \mathbf{A}_{p-1}^* \Delta \mathbf{y}_{t-p+1} + \boldsymbol{\epsilon}_t, \tag{16.52}$$

where

$$\mathbf{A}(1) = \mathbf{I}_n - \sum_{j=1}^{p} \mathbf{A}_j \quad \text{and} \quad \mathbf{A}_i^* = -\sum_{j=i+1}^{p} \mathbf{A}_j \quad \text{for} \quad i = 1, 2, \cdots, p-1. \tag{16.53}$$

This reparameterization is convenient because $-\mathbf{A}(1)$ summarizes the long-run properties of the series. We assume that there exist r linearly independent cointegrating vectors, so that $\boldsymbol{\beta}'\mathbf{y}_{t-1}$ is stationary, where $\boldsymbol{\beta}'$ is a $r \times n$ matrix of real numbers whose rows are linearly independent cointegrating vectors. Then $-\mathbf{A}(1) = \boldsymbol{\alpha}\boldsymbol{\beta}'$ for an $n \times r$ matrix of real numbers, $\boldsymbol{\alpha}$. Hence Equation (16.52) can be written as

$$\Delta \mathbf{y}_t = \boldsymbol{\delta}_\epsilon + \boldsymbol{\alpha}\boldsymbol{\beta}'\mathbf{y}_{t-1} + \mathbf{A}_1^* \Delta \mathbf{y}_{t-1} + \cdots + \mathbf{A}_{p-1}^* \Delta \mathbf{y}_{t-p+1} + \boldsymbol{\epsilon}_t, \tag{16.54}$$

This representation is called an ECM.

In many applications of standard ECMs, elements in $\boldsymbol{\alpha}$ are given structural interpretations as parameters of the speed of adjustment toward the long-run equilibrium represented by $\boldsymbol{\beta}'\mathbf{y}_{t-1}$. It is of interest to study conditions under which the elements in $\boldsymbol{\alpha}$ can be given such a structural interpretation. In the model of the next section, the domestic price level gradually adjusts to its PPP level with a speed of adjustment parameter b. We will investigate conditions under which b can be estimated as an element in $\boldsymbol{\alpha}$ from (16.54).

The standard ECM, (16.54), is a reduced form model. A class of structural models can be written in the following form of a structural ECM:

$$\mathbf{B}_0 \Delta \mathbf{y}_t = \boldsymbol{\mu}^* + \boldsymbol{\alpha}^* \boldsymbol{\beta}'\mathbf{y}_{t-1} + \mathbf{B}_1 \Delta \mathbf{y}_{t-1} + \cdots + \mathbf{B}_{p-1} \Delta \mathbf{y}_{t-p+1} + \mathbf{e}_t, \tag{16.55}$$

where $\mathbf{B}_i$ is an $n \times n$ matrix, $\boldsymbol{\mu}^*$ is an $n \times 1$ vector, and $\boldsymbol{\alpha}^*$ is an $n \times r$ matrix of real numbers. Here $\mathbf{B}_0$ is a nonsingular matrix of real numbers with ones along its principal diagonal, and $\mathbf{e}_t$ is a stationary n-dimensional vector of random variables with $\hat{E}[\mathbf{e}_t|\mathrm{H}_{t-\tau}] = 0$, where $\tau > 0$. Even though cointegrating vectors are not unique, we assume that there is a normalization that uniquely determines $\boldsymbol{\beta}$, so that parameters in $\boldsymbol{\alpha}^*$ have structural meanings.

In order to see the relationship between the standard ECM and the structural ECM, we premultiply both sides of (16.55) by $\mathbf{B}_0^{-1}$ to obtain the standard ECM (16.54), where $\boldsymbol{\delta}_\epsilon = \mathbf{B}_0^{-1}\boldsymbol{\mu}^*, \boldsymbol{\alpha} = \mathbf{B}_0^{-1}\boldsymbol{\alpha}^*, \mathbf{A}_i^* = \mathbf{B}_0^{-1}\mathbf{B}_i$, and $\boldsymbol{\epsilon}_t = \mathbf{B}_0^{-1}\mathbf{e}_t$. Thus the standard ECM estimated by Engle and Granger's two step method or Johansen's (1988) Maximum Likelihood method is a reduced form model. Hence it cannot be used to recover structural parameters in $\boldsymbol{\alpha}^*$, nor can the impulse-response functions based on $\boldsymbol{\epsilon}_t$ be interpreted in a structural way unless some restrictions are imposed on $\mathbf{B}_0$.

As in a VAR, various restrictions are possible for $\mathbf{B}_0$. One example is to assume that $\mathbf{B}_0$ is lower triangular. If $\mathbf{B}_0$ is lower triangular, then the first row of $\boldsymbol{\alpha}$ is equal to the first row of $\boldsymbol{\alpha}^*$, and structural parameters in the first row of $\boldsymbol{\alpha}^*$ are estimated by the standard methods to estimate an ECM.

16.5 An Exchange Rate Model with Sticky Prices

This section presents a simple exchange rate model in which the domestic price adjusts slowly toward the long-run equilibrium level implied by Purchasing Power Parity (PPP). Kim, Ogaki, and Yang (2007) use this model to motivate a particular form of a structural ECM in the previous section. This model's two main components are a slow adjustment equation and a rational expectations equation for the exchange rate. The single equation method is only based on the slow adjustment equation. The system method utilizes both the slow adjustment and rational expectations equations. A similar method was applied to an exchange rate model with the Taylor rule by Kim and Ogaki (2009).

Let p_t (p_t^*) be the log domestic (foreign) price level, and e_t be the log nominal exchange rate. We assume that these variables are first difference stationary and PPP holds in the long-run, so that the real exchange rate, $p_t - p_t^* - e_t$, is stationary, or $\mathbf{y}_t = (p_t,\ e_t,\ p_t^*)'$ is cointegrated with a cointegrating vector (1, -1, -1). Let $\mu = E[p_t - p_t^* - e_t]$, then μ can be nonzero when different units are used to measure prices in the two countries.

Using Mussa's (1982) model, the domestic price is assumed to adjust slowly to the PPP level

$$\Delta p_{t+1} = b(\mu + p_t^* + e_t - p_t) + E_t[p_{t+1}^* + e_{t+1}] - (p_t^* + e_t) \tag{16.56}$$

where $\Delta x_{t+1} = x_{t+1} - x_t$ for any variable x_t, $E[\cdot\ |I_t]$ is the expectation operator conditional on I_t, the information available to the economic agents at time t, and a positive constant b ($0 \leq b \leq 1$) is the adjustment coefficient. The idea behind (3) is that the domestic price slowly adjusts toward its PPP level of $p_t^* + e_t$, while it adjusts instantaneously to the expected change in its PPP level. The adjustment speed is slow (fast) when b is close to zero (one). From (3),

$$\Delta p_{t+1} = d + b(p_t^* + e_t - p_t) + \Delta p_{t+1}^* + \Delta e_{t+1} + \varepsilon_{t+1} \tag{16.57}$$

where $d = b\mu$, $\varepsilon_{t+1} = E_t[p_{t+1}^* + e_{t+1}] - (p_{t+1}^* + e_{t+1})$. Hence ε_{t+1} is a one-period ahead forecasting error, and $E[\varepsilon_{t+1}|I_t] = 0$. (4) can be referred to as the structural gradual adjustment equation which implies a first order AR structure for the real exchange rate. To see this, let $s_t = p_t^* + e_t - p_t$ be the log real exchange rate. Then (4) implies

$$s_{t+1} = -\ d + (1-b)s_t - \varepsilon_{t+1} \tag{16.58}$$

We define the half-life of the real exchange rate as the number of periods required for a unit shock to dissipate by one half in (5). Without measurement errors, b can be estimated by OLS directly from (4). In the presence of measurement errors, IV are necessary.

Let the money demand equation and the Uncovered Interest Parity (UIP) condition be

$$m_t = \theta_m + p_t - hi_t \tag{16.59}$$

$$i_t = i_t^* + E[e_{t+1}|I_t] - e_t \tag{16.60}$$

where m_t is the log nominal money supply minus the log real national income, i_t (i_t^*) is the nominal interest rate in the domestic (foreign) country. In (6), we are assuming that the income elasticity of money is one. From (6) and (7),

$$E[e_{t+1}|I_t] - e_t = (1/h)\{\theta_m + p_t - \omega_t - hE[(p_{t+1}^* - p_t^*)|I_t]\} \tag{16.61}$$

where $\omega_t = m_t + hr_t^*$ and r_t^* is the foreign real interest rate, $r_t^* = i_t^* - E[p_{t+1}^*|I_t] + p_t^*$.

Following Mussa (1982), solving (3) and (8) as a system of stochastic difference equations

$$p_t = E[F_t|I_{t-1}] - \sum_{j=1}^{\infty}(1-b)^j\{E[F_{t-j}|I_{t-j}] - E[F_{t-j}|I_{t-j-1}]\} \tag{16.62}$$

(16.63) $$e_t = \frac{bh+1}{bh} E[F_t|I_t] - p_t^* - \frac{1}{bh} p_t$$

where $F_t = (1-\delta)\sum_{j=0}^{\infty} \delta^j \omega_{t+j}$ and $\delta = h/(1+h)$. We assume that ω_t is first difference stationary. Since δ is a positive constant that is smaller than one, this implies that F_t is also first difference stationary. From (9) and (10), $e_t + p_t^* - p_t = \frac{bh+1}{bh}\sum_{j=0}^{\infty}(1-b)^j \{E[F_{t-j}|I_{t-j}] - E[F_{t-j}|I_{t-j-1}]\}$, which means $e_t + p_t^* - p_t$ is stationary.[7]

For a structural ECM representation from the exchange rate model, we use Hansen and Sargent's (1980; 1982) formula for linear rational expectations models. From (16.63),

(16.64) $$\Delta e_{t+1} = \frac{bh+1}{bh}(1-\delta)E[\sum_{j=0}^{\infty} \delta^j \Delta\omega_{t+j+1}|I_t] - \frac{1}{bh}\Delta p_{t+1} - \Delta p_{t+1}^* + \varepsilon_{e,t+1}$$

where $\varepsilon_{e,t+1} = \frac{bh+1}{bh}\,[E(F_{t+1}|I_{t+1}) - E(F_{t+1}|I_t)]$, so that the law of iterated expectation implies $E[\varepsilon_{e,t+1}|I_t] = 0$. The system method using Hansen and Sargent's (1982) method is applicable because this equation involves a discounted sum of expected future values of $\Delta\omega_t$.

Hansen and Sargent's (1982) method can be applied to this model by projecting the conditional expectation of the discounted sum, $E[\delta^j \Delta\omega_{t+j+1}|I_t]$, onto an econometrician's information set H_t. We take the econometrician's information set at t, H_t, to be the one generated by linear functions of current and past values of Δp_t^*. For simplicity, we follow West (1987) in that we choose a single variable to generate the information set H_t. In terms of the orthogonality condition, any variable in I_t can be used for this purpose.[8] Replacing $E[\sum_{j=0}^{\infty} \delta^j \Delta\omega_{t+j+1}|I_t]$ by the econometrician's linear forecast based on H_t in (11), we obtain

(16.65) $$\Delta e_{t+1} = \frac{bh+1}{bh}(1-\delta)\widehat{E}[\sum_{j=0}^{\infty} \delta^j \Delta\omega_{t+j+1}|H_t] - \frac{1}{bh}\Delta p_{t+1} - \Delta p_{t+1}^* + u_{2,t+1}$$

where $u_{2,t+1} = \varepsilon_{e,t+1} + \frac{bh+1}{bh}(1-\delta)E[(\sum_{j=0}^{\infty}\delta^j \Delta\omega_{t+j+1}|I_t) - \widehat{E}(\sum_{j=0}^{\infty}\delta^j \Delta\omega_{t+j+1}|H_t)]$ and $\widehat{E}[u_{2,t+1}|H_t] = 0$. Following Hansen and Sargent (1980, 1982) we obtain (See appendix A.)

$$\widehat{E}[\sum_{j=0}^{\infty}\Delta\omega_{t+j+1}|H_t] = \xi_1\Delta p_t^* + \xi_2\Delta p_{t-1}^* + ... + \xi_p\Delta p_{t-p+1}^* \tag{16.66}$$

A system of four equations will be[9]:

$$\Delta p_{t+1} = d + \Delta p_{t+1}^* + \Delta e_{t+1} - b(p_t - p_t^* - e_t) + u_{1,t+1} \tag{16.67}$$

$$\Delta e_{t+1} = -\frac{1}{bh}\Delta p_{t+1} - \Delta p_{t+1}^* + \alpha\xi_1\Delta p_t^* + \alpha\xi_2\Delta p_{t-1}^* + ... + \alpha\xi_p\Delta p_{t-p+1}^* + u_{2,t+1} \tag{16.68}$$

$$\Delta p_{t+1}^* = \beta_1\Delta p_t^* + \beta_2\Delta p_{t-1}^* + ... + \beta_p\Delta p_{t-p+1}^* + u_{3,t+1} \tag{16.69}$$

$$\Delta\omega_{t+1} = \gamma_1\Delta p_t^* + \gamma_2\Delta p_{t-1}^* + ... + \gamma_{p-1}\Delta p_{t-p+2}^* + u_{4,t+1} \tag{16.70}$$

where $\alpha = \frac{bh+1}{bh}(1-\delta)$ and $u_{1,t+1} = \varepsilon_{t+1}$ with a set of nonlinear restrictions imposed by (16.66),

$$\begin{aligned} \xi_0 &= \gamma(\delta)[1-\delta\beta(\delta)] \\ \xi_j &= \delta\gamma(\delta)[1-\delta\beta(\delta)]^{-1}(\beta_{j+1} + \delta\beta_{j+1} + ... + \delta^{p-j}\beta_p) + (\gamma_j + \delta\gamma_j + ... + \delta^{p-j}\gamma_p) \end{aligned} \tag{16.71}$$

for $j = 1, ..., p$. We call (16.67) the gradual adjustment equation, and (16.68)-(16.70) the Hansen and Sargent equations. Given the data for $[\Delta p_{t+1},\ \Delta e_{t+1}, \Delta p_{t+1}^*, \Delta\omega_{t+1}]'$, GMM can be applied to the system of four equations, (14)-(17).[10]

It is instructive to observe the relationship between the structural ECM and the reduced form ECM in the exchange rate model (See appendix B.). Comparing **G** and **B** shows that the speed of adjustment coefficient for the domestic price is

b in the structural model, while it is $b^2h/(bh+1)$ in the reduced form model. b in the structural form is not a deep structural parameter, unlike parameters of a production function or a utility function. However, it is clearly a parameter of interest because it determines the half-life of the real exchange rate. The reduced form speed of adjustment coefficient is a nonlinear function of b, and thus cannot be directly compared with the half-life estimates in the literature.

16.6 The System Method

Since standard methods of estimating (16.54) may not recover the structural parameters of interest in $\boldsymbol{\alpha}^*$, Kim, Ogaki, and Yang (2001) propose a system method based on GMM that does not require restrictions on $\mathbf{B}_0$.

To apply the system method to (14)-(17) of the exchange rate model, we need data for $\Delta\omega_t$, which requires knowledge of h. Even though h is unknown, a cointegrating regression can be applied to money demand if money demand is stable in the long-run, as in Stock and Watson (1993). For this purpose, we augment the model as follows:

$$m_t = \theta_m + p_t - hi_t + \zeta_{m,t} \tag{16.72}$$

where $\zeta_{m,t}$ is assumed to be stationary so that money demand is stable. By redefining m_t as $m_t - \zeta_{m,t}$, the same equations as those in section 3.2 are obtained. For the measurement of $\Delta\omega_t$, the *ex ante* foreign real interest rate can be replaced by the *ex post* value because of the Law of Iterated Expectations. Using (16.72), we obtain

$$\Delta\omega_{t+1} = \Delta p_{t+1} - h\Delta i_{t+1} + h\Delta i^*_{t+1} - h(\Delta p^*_{t+2} - \Delta p^*_{t+1}) \tag{16.73}$$

With this expression, $\Delta\omega_t$ can be measured from price and interest rate data once h is obtained, even if data for the monetary aggregate and national income are unavailable.

We have now obtained a system of four equations, (16.67)-(16.70). Because $E[u_{i,t}|I_{t-\tau}] = 0$ and $\widehat{E}[u_{i,t}|H_t] = 0$, we obtain a vector of IV $\mathbf{z}_{1,t}$ in $I_{t-\tau}$ for $u_{1,t}$ and $\mathbf{z}_{i,t}$ in H_t for $u_{i,t}$ $(i = 2, 3, 4)$.[11] Using the moment conditions $E[z_{i,t}u_{i,t}] = 0$ for $i = 1, ..., 4$ we form a GMM estimator, imposing the Hansen-Sargent restrictions and the other cross-equation restrictions implied by the model.[12] Given estimates of cointegrating vectors from the first step, this system method provides more efficient estimators than Kim's (2004) single equation method as long as the restrictions implied by the model are true.[13] The cross-equation restrictions can be tested by Wald, Likelihood Ratio (LR) type, and Lagrange Multiplier (LM) tests in the GMM framework (see Ogaki, 1993). When restrictions are nonlinear, LR and LM tests are known to be more reliable than Wald tests.

16.7 Tests for the Number of Cointegrating Vectors

Johansen's (1988; 1991) maximum likelihood (ML) estimation is based on an error correction representation:

$$\Delta \mathbf{y}_t = \boldsymbol{\delta}_\epsilon + \boldsymbol{\alpha}\boldsymbol{\beta}'\mathbf{y}_{t-1} + \mathbf{A}_1^*\Delta \mathbf{y}_{t-1} + \cdots + \mathbf{A}_{p-1}^*\Delta \mathbf{y}_{t-p+1} + \boldsymbol{\epsilon}_t, \quad (16.74)$$

where $\mathbf{y}_t$ and $\boldsymbol{\epsilon}_t$ are $n \times 1$ vectors of random variables, $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ are $n \times r$ matrices of real numbers, and $\mathbf{A}_i^*$'s are $n \times n$ matrices of real numbers. The first term $\boldsymbol{\alpha}\boldsymbol{\beta}'\mathbf{y}_{t-1}$ is called an error correction term.[12] Engle and Granger (1987) show that first difference stationary $\mathbf{y}_t$ has a possibly infinite order error correction representation with a

[12] Johansen uses an error correction term $\alpha\boldsymbol{\beta}'\mathbf{y}_{t-p}$ instead of more conventional $\alpha\boldsymbol{\beta}'\mathbf{y}_{t-1}$. However, these two representations can be shown to be equivalent.

nonzero $\boldsymbol{\alpha}$ under general regularity conditions if $\mathbf{y}_t$ is cointegrated with r linear independent cointegrating vectors. The columns of $\boldsymbol{\beta}$ are these cointegrating vectors. It should be noted that Johansen's assumption that the error correction representation of finite order can be very restrictive in some applications. For example, Gregory, Pagan, and Smith (1993) show that linear quadratic economic models with adjustment costs imply moving average terms in the error correction representation. Phillips's (1991) ML estimation method may be useful in these circumstances.

Johansen makes an additional assumption that $\boldsymbol{\epsilon}_t$ is normally distributed and derives a maximum likelihood estimator for $\boldsymbol{\beta}$. In his procedure, all parameters are jointly estimated and his estimators are asymptotically efficient. Another way to estimate an error correction representation is to use Engle and Granger's (1987) two step estimation method. In the first step, cointegrating vectors are estimated. For example, if there is only one linear independent cointegrating vector, it can be estimated by OLS. Other efficient estimators may be used in this first step. Then the rest of the parameters in the error correction representation are estimated in the second step. Since cointegrating vector estimators converge faster than $\sqrt{T}$, the first step estimation does not affect the asymptotic distributions of the second step estimators. In the second step, only stationary variables are involved, so standard econometric theory can be used. See 16.C for Johansen's maximum likelihood estimation and the cointegration rank test for detail.

Johansen's (1988; 1991) likelihood ratio tests and Stock and Watson's (1988a) tests for common trends are often used to determine the number of cointegrating vectors in a system. These tests take the null hypothesis that a $n \times 1$ vector process $\mathbf{y}_t$ has $r \geq 0$ linear independent cointegrating vectors (or it has $n - r$ common stochastic trends) against the alternative that it has $k > r$ linear independent cointegrating vectors (or it has $n - k$ common stochastic trends). Hence if $r = 0$, these statistics test the null hypothesis of no cointegration against the alternative of cointegration.

Podivinsky's (1998) Monte Carlo results suggest that there can be severe size distortion problem with Johansen's tests when the sample size is small. For example, when there is no cointegrating vector in the data generation process and when asymptotic critical values are used, he finds a tendency for the test with the null hypothesis of $r = 0$ to overreject and the test with the null hypothesis of $r \leq 1$ to underreject.

16.8 How Should an Estimation Method be Chosen?

There exist many estimation and testing methods for cointegration. It is advisable for an applied researcher to try at least two methods and check sensitivity of empirical results. When the researcher chooses a main method to be used, the following considerations naturally come to mind.

16.8.1 Are Short-Run Dynamics of Interest?

If, in addition to cointegrating vectors, the short-run dynamics are of interest, then it seems (at least conceptually) natural to estimate short-run dynamics and cointegrating vectors simultaneously. For example, this process can be done by applying Johansen's ML method to estimate an error correction model.

On the other hand, the researcher is often interested in the cointegrating vector but not in short-run dynamics (see, e.g., Atkeson and Ogaki, 1996; Clarida, 1994, 1996; Ogaki, 1992). In such cases, it is desirable to avoid making unnecessary assumptions about short-run dynamics. An estimation method that uses a nonparametric method to estimate long-run covariance parameters such as CCR is natural in these circumstances.

16.8.2 The Number of the Cointegrating Vectors

In some empirical applications, the researcher may have many economic variables and may not have any guidance from economic models about which variables may be cointegrated. In such applications, tests for the number of cointegrating vectors are useful. It should be noted, however, that these tests may not have very good small sample properties because of the near observational equivalence problem discussed in Section 13.5. For this reason, it is desirable to use economic models to give some a priori information about which variables should be cointegrated.

In some applications, an economic model implies that there exist two or more linearly independent cointegrating vectors. In this case of multiple cointegrating vectors in a cointegrating regression, neither OLS nor CCR can be used to identify cointegrating vectors. Tests for the null of cointegration based on CCR discussed above also assume that there is only one cointegrating vector and hence cannot be used. However, it is sometimes possible to use a priori information from economic models to handle multiple cointegrating vectors with the CCR methodology.[13] Johansen's ML method has an advantage that it allows multiple cointegrating vectors. However, as pointed out by Park (1990) and Pagan (1995) among others, cointegrating vectors may not be identified even by the Johansen's ML method.

[13] See Kakkar and Ogaki (1993) for an example of an empirical application.

16.8.3 Small Sample Properties

It is known that Johansen's ML estimates and test results can be very sensitive to the choice of the order of autoregression in empirical applications (see, e.g., Stock and Watson, 1993). Therefore, it is important to check sensitivity of empirical results with respect to the order of autoregression when Johansen's method is used. This sensitivity may be related to the fact that Johansen's estimator for a normalized cointegrating vector has a very large mean square error when the sample size is small (see Park and Ogaki, 1991). Gonzalo (1993) also reports this property even though he emphasizes that Johansen's estimator has good small sample properties when the sample size is increased. Podivinsky's (1998) result that Johansen's likelihood ratio tests have severe size distortion problems in some circumstances discussed in Section 16.7 may be due to these observations.

Park and Ogaki (1991) find that the CCR estimator typically has smaller mean square errors than Johansen's ML estimator when the prewhitening method is used. Han and Ogaki (1991) find that Park's tests for the null of cointegration have reasonable small sample properties.

To improve small sample properties of CCR estimators, iterations on the estimation of the long-run covariance parameters are recommended. In empirical applications of CCR, OLS is typically used as an initial estimator. Since OLS coincides with CCR when there is no correlation between the disturbance term and the first difference of the regressors at all leads and lags, the initial OLS may be called the first stage CCR. The second stage CCR is obtained from the long-run covariance parameters calculated from the first stage CCR estimates. The third stage CCR is obtained from the long-run covariance parameters calculated from the second stage CCR estimates, and so on. Park and Ogaki (1991) report that the small sample properties of the third stage CCR estimator are typically better than those of the second stage CCR estimator. On the other hand, the fourth stage CCR estimator sometimes had a significantly larger mean square error. For Park's tests for the null of cointegration to be consistent, it is necessary to bound both the eigenvalues of the VAR prewhitening coefficient matrices and the bandwidth parameter estimate. For example, while using the first order VAR for prewhitening, Han and Ogaki (1991) bound the singular values of the VAR coefficient matrix by 0.99 and the bandwidth parameter by the square root of the sample size. When the variables are cointegrated, the CCR estimators have better small sample properties without these bounds. Consequently, they recommend reporting the third stage CCR estimates without the bounds imposed and the fourth stage CCR test results with the bounds imposed.

Appendix

16.A Estimation of the Model with Long-Run Restrictions

The three variable model in KPSW highlights a real-business-model with permanent productivity shocks. Under the assumption of constant returns to scale, a production function with stochastic trends can be described as

$$y_t = \log \lambda_t + 1 - \theta k_t \tag{16.A.1}$$

$$\log \lambda_t = \mu_\lambda + \log \lambda_{t-1} + \xi_t \tag{16.A.2}$$

where y_t and k_t denote output per capita and capital stock per capita, respectively, in logarithms. Total productivity, λ_t, follows a logarithmic random walk, and ξ_t

is *iid* with mean zero and variance σ^2. Let c_t and i_t be consumption per capita and investment per capita, respectively. In the steady state, output, consumption and investment have the same growth rate of $\frac{\mu_\lambda+\xi_t}{\theta}$ which can be interpreted as a common stochastic trend. Thus, the 'great ratios', $c_t - y_t$ and $i_t - y_t$, follow stationary stochastic processes, implying y_t, c_t and i_t are cointegrated with one common trend, or equivalently, with two cointegrating relations. Therefore, there exists only one permanent innovation, v_{1t}^k that can be interpreted as a productivity shock, ξ_t. Let $\mathbf{x}_t = (y_t, c_t, i_t)'$, then $\mathbf{\Phi}(1)$ in (16.25) becomes

$$\mathbf{\Phi}(1) = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \end{bmatrix} \tag{16.A.3}$$

Since $\mathbf{\Phi}(1)$ is normalized, the first column in 16.A.3 captures the long run effects of a unit shock of v_t^1.[14] It is straightforward to estimate structural parameters following a scheme described in Section 16.3.1 where $k = 1$, $\hat{\mathbf{A}} = (1 \;\; 1 \;\; 1)'$ and $\Pi = 1$.

To incorporate nominal shocks, a six-variable model is considered in KPSW. First, money demand has the following relation

$$m_t - p_t = \beta_y y_t - \beta_R R_t + u_t \tag{16.A.4}$$

where $m_t - p_t$ is the logarithm of real balances, R_t is the nominal interest rate, and u_t is the money-demand disturbance. Second, the Fisher equation is considered to introduce nominal shocks

$$R_t = r_t + E_t \Delta p_{t+1} \tag{16.A.5}$$

where r_t is the *ex ante* real interest rate and p_t is the logarithm of the price level. Six variables $(y_t, c_t, i_t, m_t - p_t, R_t, \Delta p_t)$ follow an $I(1)$ process and exhibit cointegrating

[14] v_{1t}^k is equal to $\frac{\xi_t}{\theta}$ so that standard deviation of v_{1t}^k is equal to $\frac{\sigma}{\theta}$.

relationships. It has already been shown that there are two cointegrating relations among three variables (y_t, c_t, i_t). An additional cointegrating relationship is captured by the money demand equation in (16.A.4) provided that money-demand disturbance is stationary. Consequently, there exist three cointegrating relationships, reflecting that the system can be described by three stochastic common trends. Letting $\mathbf{x}_t = (y_t, c_t, i_t, m_t - p_t, R_t, \Delta p_t)'$, three permanent shocks consist of a real balance shock, a neutral inflation shock, and a real interest shock so that $\mathbf{A}$ is constructed as

$$\text{(16.A.6)} \qquad \mathbf{A} = \hat{\mathbf{A}}\mathbf{\Pi} = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 0 & \phi_1 \\ 1 & 0 & \phi_2 \\ \beta_y & -\beta_R & -\beta_R \\ 0 & 1 & 1 \\ 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ \pi_{21} & 1 & 0 \\ \pi_{31} & \pi_{32} & 1 \end{bmatrix}$$

KPSW assumed $\hat{\mathbf{A}}$ to be known, and constructed the parameters in $\hat{\mathbf{A}}$ by the estimates from Dynamic OLS in each cointegrating equation. It is notable that these two cointegrating relationships are used as $c - y = \phi_1(R - \Delta p)$ and $i - y = \phi_2(R - \Delta p)$ provided that the real interest rate follows a nonstationary process. This assumption implies that the 'great ratios' exhibit permanent shifts from a permanent real interest shock.[15] The issue on nonstationarity of real interest is in order. The null hypothesis that the *ex post* real interest rate[16] has a unit root is investigated using the Dickey-Fuller test, and is not rejected at the 10% significance level. This model is a benchmark in KPSW.

This property, in turn, implies that ϕ_1 and ϕ_2 are zero since regression of the $I(0)$ variable on the $I(1)$ variable gives the estimate of zero from the theoretical

[15] A higher real interest rate raises the consumption-output ratio and lowers the investment-output ratio, which implies that ϕ_1 is positive and ϕ_2 is negative.

[16] Three nominal interest rates are used in King *et al. (1989); three month U.S. Treasury bills, an average rate on four to six month commercial paper, and the yield on a portfolio of high-grade longer term corporate bonds.*

viewpoint.[17] KPSW also investigate sensitive analysis other than the benchmark model. First, the coefficients, ϕ_1 and ϕ_2, are set equal to zero. This modification, however, does not affect the main results in the benchmark model. Second, assuming that real interest rates are stationary, a model with four cointegrating relationships is considered, where two stochastic common trends are interpreted as a real balance shock and a neutral inflation shock. In this case, $\hat{\mathbf{A}}$ is constructed as

(16.A.7) $$\hat{\mathbf{A}} = \begin{bmatrix} 1 & 0 \\ 1 & 0 \\ 1 & 0 \\ \beta_y & -\beta_R \\ 0 & 1 \\ 0 & 1 \end{bmatrix}$$

The main conclusions, however, in the benchmark model are still robust after this modification.

This section explains how we can construct $\hat{\mathbf{A}}$ from the estimates of cointegrating vectors. Engle and Granger (1987) showed:

(16.A.8) $$\boldsymbol{\beta}'\boldsymbol{\Psi}(1) = \mathbf{0},$$

which by the property of cointegration implies that $\boldsymbol{\beta}'\mathbf{x}_t$ is stationary. It follows from $\boldsymbol{\Phi}(1) = \boldsymbol{\Psi}(1)\boldsymbol{\Phi}_0$ and (16.25) that

(16.A.9) $$\boldsymbol{\beta}'\mathbf{A} = \mathbf{0} \quad \text{or} \quad \boldsymbol{\beta}'\hat{\mathbf{A}} = \mathbf{0}.$$

This property enables one to choose $\hat{\mathbf{A}} = \boldsymbol{\beta}_\perp$ after re-ordering $\mathbf{x}_t$ conformably with $\boldsymbol{\beta}_\perp$, in which $\boldsymbol{\beta}_\perp$ is an $n \times k$ orthogonal matrix of cointegrating vectors, $\boldsymbol{\beta}$, satisfying $\boldsymbol{\beta}'\boldsymbol{\beta}_\perp = \mathbf{0}$. Johansen (1995) proposed a method to choose $\boldsymbol{\beta}_\perp$ by:

(16.A.10) $$\boldsymbol{\beta}_\perp = (\mathbf{I}_n - \mathbf{S}(\boldsymbol{\beta}'\mathbf{S})^{-1}\boldsymbol{\beta}')\mathbf{S}_\perp,$$

[17] ϕ_1 and ϕ_2 are estimated as 0.0033(0.0022) −0.0028(0.0050), respectively, where values in parentheses are standard errors, implying coefficients are not significantly different from zero.

where $\mathbf{S}$ is an $n \times r$ selection matrix, $(\mathbf{I}_r \;\; \mathbf{0})'$, and $\mathbf{S}_\perp$ is an $n \times k$ selection matrix, $(\mathbf{0} \;\; \mathbf{I}_k)'$. Note that $\boldsymbol{\beta}$ is identified up to the space spanned by $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$. This condition does not necessarily mean that each cointegrating vector is identified, because $\boldsymbol{\alpha}\boldsymbol{\beta}' = \boldsymbol{\alpha}\mathbf{F}\mathbf{F}^{-1}\boldsymbol{\beta}' = \tilde{\boldsymbol{\alpha}}\tilde{\boldsymbol{\beta}}'$, i.e., any linear combination of each cointegrating vector is a cointegrating vector. The model does not require the identification of each cointegrating vector. Park (1990) argues that the identification condition is not required a priori but is necessary for proper interpretation of the estimated results.

Since $\boldsymbol{\beta}_\perp$ is normalized so that the last $k \times k$ submatrix is an identity matrix, one should *re-arrange* the variables $\mathbf{x}_t$ conformably in order to maintain Blanchard and Quah (1989)-type long-run restrictions. Alternatively, one may *re-normalize* $\boldsymbol{\beta}_\perp$ as shown below. Consider the six-variable model in KPSW, for instance. Let $\mathbf{x}_t$ be $(y_t, c_t, i_t, m_t - p_t, R_t, \Delta p_t)'$, in which $m_t - p_t$ is the logarithm of the real balance, R_t is the nominal interest rate, and p_t is the logarithm of the price level. KPSW noted that there are three permanent shocks: a real balanced growth shock, a neutral inflation shock, and a real interest shock. We impose long-run restrictions that a neutral inflation shock has no long-run effect on output, and that a real interest rate shock has no long-run effect on either output or the inflation rate. These restrictions imply a specific form of $\hat{\boldsymbol{\beta}}_\perp$ as in:

$$\text{(16.A.11)} \qquad \mathbf{A} = \hat{\boldsymbol{\beta}}_\perp \boldsymbol{\Pi} = \begin{bmatrix} 1 & 0 & 0 \\ \times & \times & \times \\ \times & \times & \times \\ \times & \times & \times \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ \pi_{21} & 1 & 0 \\ \pi_{31} & \pi_{32} & 1 \end{bmatrix},$$

where $\times$ denotes that those parameters are not restricted other than $\boldsymbol{\beta}'\hat{\boldsymbol{\beta}}_\perp = 0$. From

$\mathbf{A} = \hat{\mathbf{A}}\mathbf{\Pi}$, we can choose $\hat{\mathbf{A}}$ using:[18]

$$\hat{\mathbf{A}} = \hat{\boldsymbol{\beta}}_{\perp}. \tag{16.A.12}$$

16.B Monte Carlo Integration

The literature on confidence intervals for impulse response estimates is well explained by Kilian (1998), which can be categorized by the following three traditional methods: the asymptotic interval method (see Lütkepohl, 1990), the parametric Monte Carlo integration method (see Doan, 1992; Sims and Zha, 1999), and the nonparametric bootstrap interval method (see Runkle, 1987). We provide the Monte Carlo integration method used in KPSW.[19]

It is convenient to rewrite the reduce-form VECM in (16.17) as:

$$\begin{aligned} \Delta\mathbf{x}_t' &= \boldsymbol{\delta}_\epsilon' + \mathbf{x}_{t-1}'\boldsymbol{\beta}\boldsymbol{\alpha}' + \sum_{i=1}^{p-1}\Delta\mathbf{x}_{t-i}'\mathbf{A}_i^{*\prime} + \boldsymbol{\epsilon}_t' \\ &= \mathbf{X}_t'\boldsymbol{\theta} + \boldsymbol{\epsilon}_t' \end{aligned} \tag{16.B.13}$$

where $\mathbf{X}_t' = (1, \mathbf{x}_{t-1}'\boldsymbol{\beta}, \Delta\mathbf{x}_{t-1}', \cdots, \Delta\mathbf{x}_{t-p+1}')$, and $\boldsymbol{\theta}' = (\boldsymbol{\delta}_\epsilon, \boldsymbol{\alpha}, \mathbf{A}_1^*, \cdots, \mathbf{A}_{p-1}^*)$. Stacking (16.B.13) for $t = 1, \cdots, T$, the model is represented by the following matrix form:

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\theta} + \mathbf{U} \tag{16.B.14}$$

Assuming that u_t is i.i.d. and normally distributed, Zellner (1971) finds that $\mathbf{\Sigma}$ follows the Normal-inverse Wishart posterior distribution, with the prior, $f(vec(\boldsymbol{\theta}), \mathbf{\Sigma}) \sim |\mathbf{\Sigma}|^{-\frac{n+1}{2}}$:

$$\mathbf{\Sigma}^{-1} \sim Wishart((T\mathbf{\Sigma}_0))^{-1}, T) \quad \text{with given} \quad \mathbf{\Sigma}_0, \tag{16.B.15}$$

[18] KPSW, instead, assume that $\hat{\mathbf{A}}$ is known *a priori*, which is estimated by dynamic OLS in each cointegrating equation.

[19] Kilian (1998) examines the accuracy of these confidence intervals in the small samples, and proposes the bootstrap-after-bootstrap method. He finds from Monte Carlo simulations that his method is the best, the Monte Carlo integration method is the second best, the asymptotic interval is the third, and the standard bootstrap interval method is the worst.

and

$$\boldsymbol{\theta} \sim N(\boldsymbol{\theta}_0, \boldsymbol{\Sigma} \otimes (\mathbf{X}'\mathbf{X})^{-1}), \tag{16.B.16}$$

where $\boldsymbol{\theta}_0$ and $\boldsymbol{\Sigma}_0$ are the estimates of $\boldsymbol{\theta}$ and $\boldsymbol{\Sigma}$, respectively, from OLS or MLE.

The algorithm for estimating confidence intervals of impulse responses is as follows:

1. Estimate (16.17) and let $\boldsymbol{\beta}_0$, $\boldsymbol{\theta}_0$ and $\boldsymbol{\Sigma}_0$ be these estimates.
2. Let $\mathbf{A}$ be a lower triangular matrix of Choleski decomposition of $(\mathbf{X}'\mathbf{X})^{-1}$.
3. Let $\mathbf{S}^{-1}$ be a lower triangular matrix of Choleski decomposition of $\boldsymbol{\Sigma}_0^{-1}$.
4. Generate $n \times T$ random numbers, $\mathbf{w}_b$, from the normal distribution, $N(0, \frac{1}{T})$.
5. Generate $(n(p-1)+r+1) \times n$ random numbers, $\mathbf{u}_b$, from the standard normal distribution, $N(0, 1)$.
6. Let $\mathbf{r}_b = \mathbf{w}_b'\mathbf{S}^{-1}$, and get $\boldsymbol{\Sigma}_b^{-1} = \mathbf{r}_b'\mathbf{r}_b$.
7. Let $\mathbf{S}_b$ be a lower triangular matrix of Choleski decomposition of $\boldsymbol{\Sigma}_b$.
8. Let $\boldsymbol{\theta} = \boldsymbol{\theta}_0 + \mathbf{e}_b$, in which $\mathbf{e}_b = \mathbf{A}\mathbf{u}_b\mathbf{S}_b'$. Then, $\boldsymbol{\theta} \sim N(\boldsymbol{\theta}_0, \boldsymbol{\Sigma}_b \otimes (\mathbf{X}'\mathbf{X})^{-1})$.[20]
9. Draw impulse responses, $\mathbf{ir}_b$, as described in Section 16.3.3.

[20]Note that $var(\mathbf{e}_b) = var(vec(\mathbf{e}_b)) = var((\mathbf{S}_b \otimes \mathbf{A})vec(\mathbf{u}_b)) = \mathbf{S}_b\mathbf{S}_b' \otimes \mathbf{A}\mathbf{A}' = \boldsymbol{\Sigma}_b \otimes (\mathbf{X}'\mathbf{X})^{-1}$. RATS uses $vec(\mathbf{e}_b) = (\mathbf{S}_b \otimes \mathbf{I}_{n(p-1)+r+1})vec(\mathbf{A}\mathbf{u}_b)$, which is the same as what this text uses. Note that $(\mathbf{S}_b \otimes \mathbf{A})vec(\mathbf{u}_b) = vec(\mathbf{A}\mathbf{u}_b\mathbf{S}_b') = (\mathbf{S}_b \otimes I_n)vec(\mathbf{A}\mathbf{u}_b)$, in which $vec(\mathbf{ABC}) = (\mathbf{C}' \otimes \mathbf{A})vec(\mathbf{B})$ is used for transformation.

10. Repeat 4 $\sim$ 9, B times, and calculate 95% upper and lower bands of impulse responses using[21]

$$\text{(16.B.17)} \quad Upper = \frac{1}{B}\sum_{b=1}^{B}\mathbf{ir}_b + 2(\frac{1}{B}\sum_{b=1}^{B}\mathbf{ir}_b^2 - (\frac{1}{B}\sum_{b=1}^{B}\mathbf{ir}_b)^2))^{\frac{1}{2}}$$

and

$$\text{(16.B.18)} \quad Lower = \frac{1}{B}\sum_{b=1}^{B}\mathbf{ir}_b - 2(\frac{1}{B}\sum_{b=1}^{B}\mathbf{ir}_b^2 - (\frac{1}{B}\sum_{b=1}^{B}\mathbf{ir}_b)^2))^{\frac{1}{2}}.$$

16.C Johansen's Maximum Likelihood Estimation and Cointegration Rank Tests

To see Johansen's method in detail, consider the VAR(p) model

$$\text{(16.C.19)} \qquad \mathbf{y}_t = \boldsymbol{\delta}_\epsilon + \mathbf{A}_1\mathbf{y}_{t-1} + \cdots + \mathbf{A}_p\mathbf{y}_{t-p} + \boldsymbol{\epsilon}_t,$$

where $\mathbf{y}_t$ is an $n \times 1$ vector of variables assumed to be $I(1)$. If $\mathbf{y}_t$ is cointegrated, then there exists the following VECM representation proposed by Engle and Granger (1987):

$$\text{(16.C.20)} \quad \Delta\mathbf{y}_t = \boldsymbol{\delta}_\epsilon + \boldsymbol{\alpha}\boldsymbol{\beta}'\mathbf{y}_{t-1} + \mathbf{A}_1^*\Delta\mathbf{y}_{t-1} + \cdots + \mathbf{A}_{p-1}^*\Delta\mathbf{y}_{t-p+1} + \boldsymbol{\epsilon}_t,$$

where $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ have full column rank of r, the number of cointegrating vectors.

We can concentrate on $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ from a partial regression:

$$\text{(16.C.21)} \quad \text{Regress } \Delta\mathbf{y}_t \text{ on } \mathbf{1}, \Delta\mathbf{y}_{t-1}, \cdots, \Delta\mathbf{y}_{t-p+1} \rightarrow \text{ Get residuals : } \mathbf{R}_{0t}$$

$$\text{(16.C.22)} \quad \text{Regress } \mathbf{y}_{t-1} \text{ on } \mathbf{1}, \Delta\mathbf{y}_{t-1}, \cdots, \Delta\mathbf{y}_{t-p+1} \rightarrow \text{ Get residuals : } \mathbf{R}_{kt}$$

Then, we have a concentrated regression:

$$\text{(16.C.23)} \qquad \mathbf{R}_{0t} = \boldsymbol{\alpha}\boldsymbol{\beta}'\mathbf{R}_{kt} + \boldsymbol{\epsilon}_t$$

[21] Note that we fix cointegrating vectors, $\boldsymbol{\beta}$, and generate parameters from a normal distribution, $N(\boldsymbol{\theta}_0, \boldsymbol{\Sigma}_b \otimes (\mathbf{X}'\mathbf{X})^{-1})$. Note also that we do not update $\mathbf{S}$.

For notational convenience, let

$$\mathbf{S}_{ij} = \frac{1}{T}\sum_{t=1}^{T}\mathbf{R}_{it}\mathbf{R}'_{jt}, \qquad i,j = 0,k \tag{16.C.24}$$

Note that $\boldsymbol{\alpha}$ can be easily estimated from (16.C.23) provided that $\boldsymbol{\beta}$ is known:

$$\begin{aligned} \hat{\boldsymbol{\alpha}}' &= (\boldsymbol{\beta}'\mathbf{R}'_k\mathbf{R}_k\boldsymbol{\beta})^{-1}\boldsymbol{\beta}'\mathbf{R}'_k\mathbf{R}_0 \\ &= (\boldsymbol{\beta}'\mathbf{S}_{kk}\boldsymbol{\beta})^{-1}\boldsymbol{\beta}'\mathbf{S}_{k0}. \end{aligned} \tag{16.C.25}$$

Johansen (1988) estimates $\boldsymbol{\beta}$ using MLE. Consider MLE for

$$\mathbf{Y} = XB + \mathbf{U}, \qquad u_t \sim N(\mathbf{0}, \boldsymbol{\Sigma}). \tag{16.C.26}$$

Then, the log likelihood of (16.C.26) is

$$\log L = -\frac{T}{2}\log 2\pi - \frac{T}{2}\log|\boldsymbol{\Sigma}| - \frac{1}{2}(\mathbf{Y}-\mathbf{XB})'\boldsymbol{\Sigma}^{-1}(\mathbf{Y}-\mathbf{XB}) \tag{16.C.27}$$

The FOC of (16.C.27) for $\boldsymbol{\Sigma}$ is:

$$\hat{\boldsymbol{\Sigma}} = \frac{1}{T}(\mathbf{Y}-\mathbf{XB})'(\mathbf{Y}-\mathbf{XB}) \tag{16.C.28}$$

Plug (16.C.28) in (16.C.27), then we get a concentrated likelihood:

$$\log L = \text{constant} - \frac{T}{2}\log|\hat{\boldsymbol{\Sigma}}|, \tag{16.C.29}$$

which is proportional to

$$L_{max} = |\hat{\boldsymbol{\Sigma}}|^{-\frac{T}{2}}. \tag{16.C.30}$$

Let $L(\boldsymbol{\beta}) = |\hat{\boldsymbol{\Sigma}}|^{-\frac{T}{2}}$. Then,

$$\begin{aligned} |L(\boldsymbol{\beta})|^{-\frac{2}{T}} &= |\hat{\boldsymbol{\Sigma}}| \\ &= |\frac{1}{T}(\mathbf{R}_0 - \mathbf{R}_k\boldsymbol{\beta}\boldsymbol{\alpha}')'(\mathbf{R}_0 - \mathbf{R}_k\boldsymbol{\beta}\boldsymbol{\alpha}')| \\ &= |\frac{1}{T}(\mathbf{R}_0\mathbf{R}_0 - \boldsymbol{\alpha}\boldsymbol{\beta}'\mathbf{R}'_k\mathbf{R}_k\boldsymbol{\beta}\boldsymbol{\alpha}')| \\ &= |\mathbf{S}_{00} - \mathbf{S}_{0k}\boldsymbol{\beta}(\boldsymbol{\beta}'\mathbf{S}_{kk}\boldsymbol{\beta})^{-1}\boldsymbol{\beta}'\mathbf{S}_{k0}| \end{aligned} \tag{16.C.31}$$

So,

$$\begin{aligned}(16.C.32)\qquad \max_{\beta} L(\boldsymbol{\beta}) \;&\Leftrightarrow\; \min_{\beta} |\mathbf{S}_{00} - \mathbf{S}_{0k}\boldsymbol{\beta}(\boldsymbol{\beta}'\mathbf{S}_{kk}\boldsymbol{\beta})^{-1}\boldsymbol{\beta}'\mathbf{S}_{k0}| \\ &\Leftrightarrow\; \min_{\beta} |\boldsymbol{\beta}'\mathbf{S}_{kk}\boldsymbol{\beta} - \boldsymbol{\beta}'\mathbf{S}_{k0}\mathbf{S}_{00}^{-1}\mathbf{S}_{0k}\boldsymbol{\beta}|\frac{|\mathbf{S}_{00}|}{|\boldsymbol{\beta}'\mathbf{S}_{kk}\boldsymbol{\beta}|} \\ &\Leftrightarrow\; \max_{\beta} \frac{|\boldsymbol{\beta}'\mathbf{S}_{kk}\boldsymbol{\beta}|}{|\boldsymbol{\beta}'(\mathbf{S}_{kk} - \mathbf{S}_{k0}\mathbf{S}_{00}^{-1}\mathbf{S}_{0k})\boldsymbol{\beta}|}\frac{1}{|\mathbf{S}_{00}|}\end{aligned}$$

At the second line, we use the following formula:

$$(16.C.33)\qquad \begin{vmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{C} & \mathbf{D} \end{vmatrix} = |\mathbf{A}|\;|\mathbf{D} - \mathbf{C}\mathbf{A}^{-1}\mathbf{B}| = |\mathbf{D}|\;|\mathbf{A} - \mathbf{B}\mathbf{D}^{-1}\mathbf{C}|$$

Thus,

$$(16.C.34)\qquad |\mathbf{A} - \mathbf{B}\mathbf{D}^{-1}\mathbf{C}| = |\mathbf{D} - \mathbf{C}\mathbf{A}^{-1}\mathbf{B}|\frac{|\mathbf{A}|}{|\mathbf{D}|},$$

where $\mathbf{A} = \mathbf{S}_{00}, \mathbf{B} = \mathbf{S}_{0k}\boldsymbol{\beta}, \mathbf{C} = \boldsymbol{\beta}'\mathbf{S}_{k0}$, and $\mathbf{D} = \boldsymbol{\beta}'\mathbf{S}_{kk}\boldsymbol{\beta}$. Note also that FOC for

$$(16.C.35)\qquad \max_{\mathbf{x}} \frac{\mathbf{x}'\mathbf{A}\mathbf{x}}{\mathbf{x}'\mathbf{B}\mathbf{x}} \quad (\equiv \lambda)$$

is

$$(16.C.36)\qquad (\mathbf{A} - \lambda\mathbf{B})\mathbf{x} = \mathbf{0},$$

where λ is an eigenvalue, and $\mathbf{x}$ is an eigenvector. Therefore, (16.C.32) becomes an eigenvalue problem. Let

$$(16.C.37)\qquad \lambda_0 = \max_{\beta} \frac{|\boldsymbol{\beta}'\mathbf{S}_{kk}\boldsymbol{\beta}|}{|\boldsymbol{\beta}'(\mathbf{S}_{kk} - \mathbf{S}_{k0}\mathbf{S}_{00}^{-1}\mathbf{S}_{0k})\boldsymbol{\beta}|}.$$

Then, the FOC is

$$\begin{aligned}(16.C.38)\qquad &(\mathbf{S}_{kk} - \lambda_0(\mathbf{S}_{kk} - \mathbf{S}_{k0}\mathbf{S}_{00}^{-1}\mathbf{S}_{0k}))\boldsymbol{\beta} = \mathbf{0} \\ \Leftrightarrow\;& ((1-\lambda_0)\mathbf{S}_{kk} + \lambda_0(\mathbf{S}_{k0}\mathbf{S}_{00}^{-1}\mathbf{S}_{0k}))\boldsymbol{\beta} = \mathbf{0} \\ \Leftrightarrow\;& (\lambda_0(\mathbf{S}_{k0}\mathbf{S}_{00}^{-1}\mathbf{S}_{0k}) - (\lambda_0 - 1)\mathbf{S}_{kk})\boldsymbol{\beta} = \mathbf{0} \\ \Leftrightarrow\;& (\mathbf{S}_{k0}\mathbf{S}_{00}^{-1}\mathbf{S}_{0k} - (1 - \frac{1}{\lambda_0})\mathbf{S}_{kk})\boldsymbol{\beta} = \mathbf{0} \\ \Leftrightarrow\;& (\mathbf{S}_{k0}\mathbf{S}_{00}^{-1}\mathbf{S}_{0k} - \lambda\mathbf{S}_{kk})\boldsymbol{\beta} = \mathbf{0},\end{aligned}$$

where $\lambda = 1 - \frac{1}{\lambda_0}$. Note that λ and $\boldsymbol{\beta}$ are an eigenvalue and an eigenvector of $\mathbf{S}_{kk}^{-1}\mathbf{S}_{k0}\mathbf{S}_{00}^{-1}\mathbf{S}_{0k}$, respectively. Therefore, our maximization problem is reduced to find an eigenvalue and eigenvector of $\mathbf{S}_{kk}^{-1}\mathbf{S}_{k0}\mathbf{S}_{00}^{-1}\mathbf{S}_{0k}$.

Having estimated the model, we can construct the cointegration rank tests as follows. From (16.C.30), (16.C.32) and (16.C.37), we get

$$|L_{max}(\boldsymbol{\beta})|^{-\frac{2}{T}} = |\mathbf{S}_{00}| \prod_{i=1}^{r} \frac{1}{\lambda_{0i}} \tag{16.C.39}$$

$$L_{max}(\boldsymbol{\beta}) = -\frac{T}{2}|\mathbf{S}_{00}| \prod_{i=1}^{r}(1-\lambda_i) \tag{16.C.40}$$

Therefore, we get the LR test (or Trace test) as:

$$\begin{aligned} LR &= -2\log\frac{L_{max}(H_0 = r)}{L_{max}(H_1 = n)} \\ &= -T\sum_{i=r+1}^{n}\log(1-\lambda_i) \end{aligned} \tag{16.C.41}$$

and the maximum eigenvalue test (or λ_{max} test) as:

$$\begin{aligned} \lambda_{max} &= -2\log\frac{L_{max}(H_0 = r)}{L_{max}(H_1 = r+1)} \\ &= -T\log(1-\lambda_{r+1}). \end{aligned} \tag{16.C.42}$$

Note that the alternative hypothesis is different in each test. For large values of test statistics, we reject the null hypothesis that there exist r cointegrating vectors, $H_0 = r$. Johansen (1995) gives the critical values, and Osterwald-Lenum (1992) provides revised critical values.

Johansen (1995) considers five models with respect to data properties as well as cointegrating relations as follows: i) a model with a quadratic trend in $\mathbf{y}_t$ (hflag=1):

$$\Delta\mathbf{y}_t = \boldsymbol{\delta}_\epsilon + \boldsymbol{\rho}_0 t + \boldsymbol{\alpha}\boldsymbol{\beta}'\mathbf{y}_{t-1} + \mathbf{A}_1^*\Delta\mathbf{y}_{t-1} + \cdots + \mathbf{A}_{p-1}^*\Delta\mathbf{y}_{t-p+1} + \boldsymbol{\epsilon}_t, \tag{16.C.43}$$

ii) a model with a linear trend in $\mathbf{y}_t$ (hflag=2), in which deterministic cointegration is not satisfied:

$$\Delta \mathbf{y}_t = \boldsymbol{\delta}_\epsilon + \boldsymbol{\rho}_0 t + \boldsymbol{\alpha}\boldsymbol{\beta}'\mathbf{y}_{t-1} + \mathbf{A}_1^*\Delta\mathbf{y}_{t-1} + \cdots + \mathbf{A}_{p-1}^*\Delta\mathbf{y}_{t-p+1} + \boldsymbol{\epsilon}_t, \tag{16.C.44}$$

iii) a model with a linear trend in $\mathbf{y}_t$ (hflag=3), in which deterministic cointegration is satisfied (cotrended):

$$\Delta \mathbf{y}_t = \boldsymbol{\delta}_\epsilon + \boldsymbol{\alpha}(\boldsymbol{\beta}'\mathbf{y}_{t-1} + \boldsymbol{\rho}_1 t) + \mathbf{A}_1^*\Delta\mathbf{y}_{t-1} + \cdots + \mathbf{A}_{p-1}^*\Delta\mathbf{y}_{t-p+1} + \boldsymbol{\epsilon}_t, \tag{16.C.45}$$

iv) a model with no trend in $\mathbf{y}_t$ (hflag=4):

$$\Delta \mathbf{y}_t = \boldsymbol{\alpha}(\boldsymbol{\beta}'\mathbf{y}_{t-1} + \boldsymbol{\rho}_0) + \mathbf{A}_1^*\Delta\mathbf{y}_{t-1} + \cdots + \mathbf{A}_{p-1}^*\Delta\mathbf{y}_{t-p+1} + \boldsymbol{\epsilon}_t, \tag{16.C.46}$$

and v) a model with no trend in $\mathbf{y}_t$ (hflag=5):

$$\Delta \mathbf{y}_t = \boldsymbol{\alpha}\boldsymbol{\beta}'\mathbf{y}_{t-1} + \mathbf{A}_1^*\Delta\mathbf{y}_{t-1} + \cdots + \mathbf{A}_{p-1}^*\Delta\mathbf{y}_{t-p+1} + \boldsymbol{\epsilon}_t. \tag{16.C.47}$$

Johansen (1995) illustrates how to estimate restricted cointegrating vectors. Consider a trivariate model with two cointegrating vectors. Let $\mathbf{y}_t = (\mathbf{y}_{1t}, \mathbf{y}_{2t}, \mathbf{y}_{3t})'$ and $\boldsymbol{\beta} = [\boldsymbol{\beta}_1|\boldsymbol{\beta}_2]$. One may impose a restriction of $\boldsymbol{\beta}_{11} = \boldsymbol{\beta}_{13}$ using $\mathbf{H}_1\boldsymbol{\varphi}_1 = \boldsymbol{\beta}_1$ and $\mathbf{H}_2\boldsymbol{\varphi}_2 = \boldsymbol{\beta}_2$, where $\mathbf{H}_i$ is an $n \times (n - q_i)$ matrix, $\boldsymbol{\varphi}_i$ is an $(n - q_i) \times 1$ matrix, and q_i is the number of restrictions on each cointegrating vector. In this particular example, letting

$$\mathbf{H}_1 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ -1 & 0 \end{bmatrix}, \quad \mathbf{H}_2 = \mathbf{I}_3 \tag{16.C.48}$$

gives the following restrictions:

$$\mathbf{H}_1\boldsymbol{\varphi}_1 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ -1 & 0 \end{bmatrix} \begin{bmatrix} \boldsymbol{\varphi}_{11} \\ \boldsymbol{\varphi}_{12} \end{bmatrix} = \begin{bmatrix} \boldsymbol{\varphi}_{11} \\ \boldsymbol{\varphi}_{12} \\ -\boldsymbol{\varphi}_{11} \end{bmatrix} = \begin{bmatrix} \boldsymbol{\beta}_{11} \\ \boldsymbol{\beta}_{12} \\ \boldsymbol{\beta}_{13} \end{bmatrix}. \tag{16.C.49}$$

References

Atkeson, A., and M. Ogaki (1996): "Wealth-Varying Intertemporal Elasticities of Substitution: Evidence from Panel and Aggregate Data," *Journal of Monetary Economics*, 38, 507–534.

Blanchard, O. J., and D. Quah (1989): "The Dynamic Effects of Aggregate Demand and Supply Disturbances," *American Economic Review*, 79(4), 655–673.

Campbell, J. Y., and R. J. Shiller (1988): "Interpreting Cointegrated Models," *Journal of Economic Dynamics and Control*, 12, 505–522.

Clarida, R. H. (1994): "Cointegration, Aggregate Consumption, and the Demand for Imports: A Structural Econometric Investigation," *American Economic Review*, 84(1), 298–308.

——— (1996): "Consumption, Import Prices, and the Demand for Imported Consumer Durables: A Structural Econometric Investigation," *Review of Economics and Statistics*, 78, 369–374.

Doan, T. A. (1992): *RATS User's Manual, Version 4*. Estima, Evanston, IL.

Engle, R. F., and C. Granger (1987): "Co-Integration and Error Correction: Representation, Estimation, and Testing," *Econometrica*, 55, 251–276.

Engle, R. F., and B. S. Yoo (1987): "Forecasting and Testing in Co-Integrated Systems," *Journal of Econometrics*, 35(1), 143–159.

Fisher, L. A., P. L. Fackler, and D. Orden (1995): "Long-run Identifying Restrictions for an Error-Correction Model of New Zealand Money, Prices and Output," *Journal of International Money and Finance*, 14(1), 127–147.

Galí, J. (1992): "How Well Does the IS-LM Model Fit Postwar U.S. Data?," *Quarterly Journal of Economics*, 107(2), 709–738.

——— (1999): "Technology, Employment, and the Business Cycle: Do Technology Shocks Explain Aggregate Fluctuations?," *American Economic Review*, 89(1), 249–271.

Gonzalo, J. (1993): "Cointegration and Aggregation," *Ricerche Economiche*, 47(3), 281–291.

Gregory, A. W., A. R. Pagan, and G. Smith (1993): "Estimating Linear Quadratic Models with Integrated Processes," in *Models, Methods and Applications of Econometrics*, ed. by P. C. B. Phillips, pp. 220–239. Basil Blackwell, Oxford.

Hamilton, J. D. (1994): *Time Series Analysis*. Princeton University Press, Princeton.

Han, H.-L., and M. Ogaki (1991): "Consumption, Income, and Cointegration: Further Analysis," RCER Working Papers No. 305.

Hansen, L. P., and T. J. Sargent (1980): "Formulating and Estimating Dynamic Linear Rational Expectations Models," *Journal of Economic Dynamics and Control*, 2(1), 7–46.

——— (1982): "Instrumental Variables Procedures for Estimating Linear Rational Expectations Models," *Journal of Monetary Economics*, 9(3), 263–296.

Jang, K. (2001a): "Impulse Response Analysis with Long Run Restrictions on Error Correction Models," Working Paper No. 01-04, Department of Economics, Ohio State University.

——— (2001b): "A Structural Vector Error Correction Model with Short-Run and Long-Run Restrictions," Manuscript.

Jang, K., and M. Ogaki (2001): "The Effects of Monetary Policy Shocks on Exchange Rates: A Structural Vector Error Correction Model Approach," Working Paper No. 01-02, Department of Economics, Ohio State University.

Johansen, S. (1988): "Statistical Analysis of Cointegration Vectors," *Journal of Economic Dynamics and Control*, 12, 231–254.

——— (1991): "Estimation and Hypothesis Testing of Cointegration Vectors in Gaussian Vector Autoregressive Models," *Econometrica*, 59(6), 1551–1580.

——— (1995): *Likelihood-Based Inference in Cointegrated Vector Autoregressive Models*. Oxford University Press, Oxford.

Kakkar, V., and M. Ogaki (1993): "Real Exchange Rates and Nontradables," Manuscript.

Kilian, L. (1998): "Small-Sample Confidence Intervals for Impulse Response Functions," *Review of Economics and Statistics*, 80(2), 218–230.

Kim, H., and M. Ogaki (2009): "Purchasing Power Parity and the Taylor Rule," Working Paper No. 09-03, Department of Economics, Ohio State University.

Kim, J. (2004): "Half-Lives of Deviations from PPP: Contrasting Traded and Nontraded Components of Consumption Baskets," *Review of International Economics*, 12(1), 162–168.

Kim, J., M. Ogaki, and M.-S. Yang (2001): "Structural Error Correction Models: Instrumental Variables Methods and Application to an Exchange Rate Model," Working Paper No. 01-01, Department of Economics, Ohio State University.

——— (2007): "Structural Error Correction Models: A System Method for Linear Rational Expectations Models and an Application to an Exchange Rate Model," *Journal of Money, Credit, and Banking*, 39(8), 2057–2075.

King, R. G., C. I. Plosser, J. H. Stock, and M. W. Watson (1989): "Stochastic Trends and Economic Fluctuations," Manuscript.

——— (1991): "Stochastic Trends and Economic Fluctuations," *American Economic Review*, 81(4), 810–840.

Lastrapes, W. D., and G. Selgin (1995): "The Liquidity Effect: Identifying Short-Run Interest Rate Dynamics Using Long-Run Restrictions," *Journal of Macroeconomics*, 17(3), 387–404.

Lütkepohl, H. (1990): "Asymptotic Distributions of Impulse Response Functions and Forecast Error Variance Decompositions of Vector Autoregressive Models," *Review of Economics and Statistics*, 72(1), 116–125.

Lütkepohl, H., and H.-E. Reimers (1992): "Impulse response analysis of cointegrated systems," *Journal of Economic Dynamics and Control*, 16, 53–78.

Mellander, E., A. Vredin, and A. Warne (1992): "Stochastic Trends and Economic Fluctuations in a Small Open Economy," *Journal of Applied Econometrics*, 7(4), 369–394.

Mussa, M. (1982): "A Model of Exchange Rate Dynamics," *Journal of Political Economy*, 90(1), 74–104.

Ogaki, M. (1992): "Engel's Law and Cointegration," *Journal of Political Economy*, 100(5), 1027–1046.

——— (1993): "Generalized Method of Moments: Econometric Applications," in *Handbook of Statistics: Econometrics*, ed. by G. S. Maddala, C. R. Rao, and H. D. Vinod, vol. 11, chap. 17, pp. 455–488. North-Holland, Amsterdam.

Ogaki, M., and J. Y. Park (1997): "A Cointegration Approach to Estimating Preference Parameters," *Journal of Econometrics*, 82(1), 107–134.

Osterwald-Lenum, M. (1992): "A Note with Quantiles of the Asymptotic Distribution of the Maximum Likelihood Cointegration Rank Test Statistics," *Oxford Bulletin of Economics and Statistics*, 54(3), 461–471.

Pagan, A. R. (1995): "Three Econometric Methodologies: An Update," in *Surveys in Econometrics*, ed. by L. T. Oxely, C. J. Roberts, D. A. R. George, and S. T. Sayer, pp. 30–41. Basil Blackwell.

Pagan, A. R., and J. C. Robertson (1995): "Structural Models of the Liquidity Effect," Manuscript.

——— (1998): "Structural Models of the Liquidity Effect," *Review of Economics and Statistics*, 80, 202–217.

Park, J. Y. (1990): "Maximum Likelihood Estimation of Simultaneous Cointegrated Models," Manuscript, Cornell University.

Park, J. Y., and M. Ogaki (1991): "VAR Prewhitening to Estimate Short-Run Dynamics: On Improved Method of Inference in Cointegrated Models," RCER Working Paper No. 281.

Phillips, P. C. B. (1991): "Optimal Inference in Cointegrated Systems," *Econometrica*, 59, 283–306.

Podivinsky, J. M. (1998): "Testing Misspecified Cointegrating Relationships," *Economics Letters*, 60(1), 1–9.

Qureshi, H. (2008): "Explosive Roots in Level Vector Autoregressive Models," Working Papers 08-02, Ohio State University, Department of Economics.

Runkle, D. E. (1987): "Vector Autoregressions and Reality," *Journal of Business and Economic Statistics*, 5, 437–442.

Shapiro, M. D., and M. W. Watson (1988): "Sources of Business Cycle Fluctuations (with Comments)," in *NBER Macroeconomic Annual*, ed. by S. Fischer, vol. 3, pp. 111–156. Cambridge: MIT Press.

Sims, C. A. (1980): "Macroeconomics and Reality," *Econometrica*, 48, 1–48.

Sims, C. A., and T. Zha (1999): "Error Bands for Impulse Responses," *Econometrica*, 67(5), 1113–1156.

STOCK, J. H., AND M. W. WATSON (1988a): "Testing for Common Trends," *Journal of the American Statistical Association*, 83(404), 1097–1107.

——— (1988b): "Variable Trends in Economic Time Series," *Journal of Economic Perspectives*, 2(3), 147–74.

——— (1993): "A Simple Estimator of Cointegrating Vectors in Higher Order Integrated Systems," *Econometrica*, 61(4), 783–820.

WARNE, A. (1991): "A Common Trends Model: Identification, Estimation and Asymptotics," Manuscript.

WEST, K. D. (1987): "A Specification Test for Speculative Bubbles," *Quarterly Journal of Economics*, 102, 553–580.

ZELLNER, A. (1971): *An Introduction to Bayesian Inference in Econometrics.* Wiley, New York.