Chapter 5

Analysis of Multiple Time Series

The alternative reference for the material in this chapter is Enders (2004) (chapters 5 and 6). Chapters 10-11 and 18-19 in Hamilton (1994) provide a more technical treatment of the material.

> Multivariate time-series analysis extends many of the ideas of univariate time-series analysis to systems of equations. The primary model used in multivariate time-series analysis is the vector autoregression (VAR). Many properties of autoregressive processes extend naturally to multivariate time-series using a slight change in notation and results from linear algebra. This chapter examines the properties of vector time-series models, estimation and identification and introduces two new concepts: Granger Causality and the Impulse Response Function. The chapter concludes by examining models of contemporaneous relationships between two or more time-series in the framework of cointegration, spurious regression and cross-sectional regression of stationary time-series.

In many applications, analyzing a time-series in isolation is a reasonable choice; in others, univariate analysis is insufficient to capture the complex dynamics among interrelated time series. For example, Campbell (1996) links financially interesting variables, including stock returns and the default premium, in a multivariate system where shocks to one variable propagate to the others. The vector autoregression (VAR) is the standard model used to model multiple *stationary* time-series. If the time series are not stationary, a different type of analysis, cointegration, is used.

5.1   Vector Autoregressions

Vector autoregressions are remarkably similar to univariate autoregressions, and most results carry over by replacing scalars with matrices and scalar operations with their linear algebra equivalent.

5.1.1   Definition

The definition of a vector autoregression is nearly identical to that of a univariate autoregression.

Definition 5.1 (Vector Autoregression of Order P). A Pth order vector autoregression, written VAR(P), is a process that evolves according to

$$\mathbf{Y}_t = \mathbf{\Phi}_0 + \mathbf{\Phi}_1 \mathbf{Y}_{t-1} + \mathbf{\Phi}_2 \mathbf{Y}_{t-2} + \ldots + \mathbf{\Phi}_P \mathbf{Y}_{t-P} + \boldsymbol{\varepsilon}_t \tag{5.1}$$

where $\mathbf{Y}_t$ is a k by 1 vector stochastic process, Φ_0 is a k by 1 vector of intercept parameters, Φ_j, $j = 1,\ldots,P$ are k by k parameter matrices and ε_t is a vector white noise process with the additional assumption that $\mathrm{E}_{t-1}[\varepsilon_t] = \mathbf{0}$.

A VAR(P) reduces to an AR(P) when $k = 1$ so that $\mathbf{Y}_t$ and the coefficient matrices, Φ_j, are scalars. A vector white noise process extends the three properties of a univariate white noise process to a vector; it is mean zero, has finite covariance and is uncorrelated with its past. The components of a vector white noise process are not assumed to be *contemporaneously* uncorrelated.

Definition 5.2 (Vector White Noise Process). A k by 1 vector-valued stochastic process, $\{\varepsilon_t\}$ is a vector white noise if

$$
\begin{aligned}
\mathrm{E}[\varepsilon_t] &= \mathbf{0}_k \\
\mathrm{E}[\varepsilon_t \varepsilon_{t-s}'] &= \mathbf{0}_{k \times k} \\
\mathrm{E}[\varepsilon_t \varepsilon_t'] &= \Sigma
\end{aligned}
\tag{5.2}
$$

for all t where Σ is a finite positive definite matrix.

The simplest VAR is a first-order bivariate specification which is equivalently expressed as

$$ \mathbf{Y}_t = \Phi_0 + \Phi_1 \mathbf{Y}_{t-1} + \varepsilon_t, $$

$$
\begin{bmatrix} Y_{1,t} \\ Y_{2,t} \end{bmatrix} = \begin{bmatrix} \phi_{1,0} \\ \phi_{2,0} \end{bmatrix} + \begin{bmatrix} \phi_{11,1} & \phi_{12,1} \\ \phi_{21,1} & \phi_{22,1} \end{bmatrix} \begin{bmatrix} Y_{1,t-1} \\ Y_{2,t-1} \end{bmatrix} + \begin{bmatrix} \varepsilon_{1,t} \\ \varepsilon_{2,t} \end{bmatrix},
$$

or

$$
\begin{aligned}
Y_{1,t} &= \phi_{1,0} + \phi_{11,1} Y_{1,t-1} + \phi_{12,1} Y_{2,t-1} + \varepsilon_{1,t} \\
Y_{2,t} &= \phi_{2,0} + \phi_{21,1} Y_{1,t-1} + \phi_{22,1} Y_{2,t-1} + \varepsilon_{2,t}.
\end{aligned}
$$

Each element of $\mathbf{Y}_t$ is a function of each element of $\mathbf{Y}_{t-1}$.

5.1.2 Properties of a VAR(1)

The properties of the VAR(1) are straightforward to derive. Importantly, section 5.2 shows that all VAR(P)s can be rewritten as a VAR(1), and so the properties of any VAR follow from those of a first-order VAR.

5.1.2.1 Stationarity

A VAR(1), driven by vector white noise shocks,

$$ \mathbf{Y}_t = \Phi_0 + \Phi_1 \mathbf{Y}_{t-1} + \varepsilon_t $$

is covariance stationary if the eigenvalues of Φ_1 are less than 1 in modulus.[1] In the univariate case, this is statement is equivalent to the condition $|\phi_1| < 1$. Assuming the eigenvalues of Φ_1 are less than one in absolute value, backward substitution can be used to show that

---

[1] The definition of an eigenvalue is:

$$\mathbf{Y}_t = \sum_{i=0}^{\infty} \Phi_1^i \Phi_0 + \sum_{i=0}^{\infty} \Phi_1^i \varepsilon_{t-i} \tag{5.3}$$

which, applying Theorem 5.3, is equivalent to

$$\mathbf{Y}_t = (\mathbf{I}_k - \Phi_1)^{-1} \Phi_0 + \sum_{i=0}^{\infty} \Phi_1^i \varepsilon_{t-i} \tag{5.4}$$

where the eigenvalue condition ensures that Φ_1^i converges to zero as i grows large.

5.1.2.2 Mean

Taking expectations of $\mathbf{Y}_t$ expressed in the backward substitution form yields

$$\begin{aligned}
\mathrm{E}\left[\mathbf{Y}_t\right] &= \mathrm{E}\left[(\mathbf{I}_k - \Phi_1)^{-1} \Phi_0\right] + \mathrm{E}\left[\sum_{i=0}^{\infty} \Phi_1^i \varepsilon_{t-i}\right] \\
&= (\mathbf{I}_k - \Phi_1)^{-1} \Phi_0 + \sum_{i=0}^{\infty} \Phi_1^i \mathrm{E}\left[\varepsilon_{t-i}\right] \\
&= (\mathbf{I}_k - \Phi_1)^{-1} \Phi_0 + \sum_{i=0}^{\infty} \Phi_1^i \mathbf{0} \\
&= (\mathbf{I}_k - \Phi_1)^{-1} \Phi_0
\end{aligned} \tag{5.5}$$

The mean of a VAR process resembles that of a univariate AR(1), $(1 - \phi_1)^{-1} \phi_0$.[2] The long-run mean depends on the intercept, Φ_0, and the inverse of Φ_1. The magnitude of the inverse is determined by

Definition 5.3 (Eigenvalue). λ is an eigenvalue of a square matrix $\mathbf{A}$ if and only if $|\mathbf{A} - \lambda \mathbf{I}_n| = 0$ where $|\cdot|$ denotes determinant.

Definition 5.4. Eigenvalues play a unique role in the matrix power operator.

Theorem 5.1 (Singular Value Decomposition). *Let $\mathbf{A}$ be an n by n real-valued matrix. Then $\mathbf{A}$ can be decomposed as $\mathbf{A} = \mathbf{U}\Lambda\mathbf{V}'$ where $\mathbf{V}'\mathbf{U} = \mathbf{U}'\mathbf{V} = \mathbf{I}_n$ and Λ is a diagonal matrix containing the eigenvales of $\mathbf{A}$.*

Theorem 5.2 (Matrix Power). *Let $\mathbf{A}$ be an n by n real-valued matrix. Then $\mathbf{A}^m = \mathbf{A}\mathbf{A}\ldots\mathbf{A} = \mathbf{U}\Lambda\mathbf{V}'\mathbf{U}\Lambda\mathbf{V}'\ldots\mathbf{U}\Lambda\mathbf{V}' = \mathbf{U}\Lambda^m\mathbf{V}'$ where Λ^m is a diagonal matrix containing each eigenvalue of $\mathbf{A}$ raised to the power m.*

The essential properties of eigenvalues for applications to VARs are given in the following theorem:

Theorem 5.3 (Convergent Matrices). *Let $\mathbf{A}$ be an n by n matrix. Then the following statements are equivalent*

- *$\mathbf{A}^m \to 0$ as $m \to \infty$.*
- *All eigenvalues of $\mathbf{A}$, λ_i, $i = 1, 2, \ldots, n$, are less than 1 in modulus ($|\lambda_i| < 1$).*
- *The series $\sum_{i=0}^{m} \mathbf{A}^m = \mathbf{I}_n + \mathbf{A} + \mathbf{A}^2 + \ldots + \mathbf{A}^m \to (\mathbf{I}_n - \mathbf{A})^{-1}$ as $m \to \infty$.*

Note: Replacing $\mathbf{A}$ with a scalar a produces many familiar results: $a^m \to 0$ as $m \to \infty$ (property 1) and $\sum_{i=0}^{m} a^m \to (1-a)^{-1}$ as $m \to \infty$ (property 3) as long as $|a| < 1$ (property 2).

[2]When a is a scalar where $|a| < 1$, then $\sum_{i=0}^{\infty} a^i = 1/(1-a)$. This result extends to a $k \times k$ square matrix $\mathbf{A}$ when all of the eigenvalues of $\mathbf{A}$ are less than 1, so that $\sum_{i=0}^{\infty} \mathbf{A}^i = (\mathbf{I}_k - \mathbf{A})^{-1}$.

the eigenvalues of Φ_1, and if any eigenvalue is close to one, then $(\mathbf{I}_k - \Phi_1)^{-1}$ is large in magnitude and, all things equal, the unconditional mean is larger. Similarly, if $\Phi_1 = \mathbf{0}$, then the mean is Φ_0 since $\{\mathbf{Y}_t\}$ is a constant plus white noise.

5.1.2.3 Variance

Before deriving the variance of a VAR(1), it useful to express a VAR in *deviations* form. Define $\mu = \mathrm{E}[\mathbf{Y}_t]$ to be the unconditional expectation (assumed it is finite). The deviations form of the VAR(P)

$$\mathbf{Y}_t = \Phi_0 + \Phi_1 \mathbf{Y}_{t-1} + \Phi_2 \mathbf{Y}_{t-2} + \ldots + \Phi_P \mathbf{Y}_{t-P} + \varepsilon_t$$

is

$$\mathbf{Y}_t - \mu = \Phi_1 (\mathbf{Y}_{t-1} - \mu) + \Phi_2 (\mathbf{Y}_{t-2} - \mu) + \ldots + \Phi_P (\mathbf{Y}_{t-P} - \mu) + \varepsilon_t \tag{5.6}$$
$$\tilde{\mathbf{Y}}_t = \Phi_1 \tilde{\mathbf{Y}}_{t-1} + \Phi_2 \tilde{\mathbf{Y}}_{t-2} + \ldots + \Phi_P \tilde{\mathbf{Y}}_{t-P} + \varepsilon_t.$$

The deviations form is mean $\mathbf{0}$ by construction, and so the backward substitution form in a VAR(1) is

$$\tilde{\mathbf{Y}}_t = \sum_{i=1}^{\infty} \Phi_1^i \varepsilon_{t-i}. \tag{5.7}$$

The deviations form translates the VAR from its original mean, μ, to a mean of $\mathbf{0}$. The process written in deviations form has the same dynamics and shocks, and so can be used to derive the long-run covariance and autocovariances and to simplify multistep forecasting. The long-run covariance is derived using the backward substitution form so that

$$\mathrm{E}\left[(\mathbf{Y}_t - \mu)(\mathbf{Y}_t - \mu)'\right] = \mathrm{E}\left[\tilde{\mathbf{Y}}_t \tilde{\mathbf{Y}}_t'\right] = \mathrm{E}\left[\left(\sum_{i=0}^{\infty} \Phi_1^i \varepsilon_{t-i}\right)\left(\sum_{i=0}^{\infty} \Phi_1^i \varepsilon_{t-i}\right)'\right] \tag{5.8}$$

$$= \mathrm{E}\left[\sum_{i=0}^{\infty} \Phi_1^i \varepsilon_{t-i} \varepsilon_{t-i}' (\Phi_1')'\right] \quad (\text{Since } \varepsilon_t \text{ is WN})$$

$$= \sum_{i=0}^{\infty} \Phi_1^i \mathrm{E}\left[\varepsilon_{t-i} \varepsilon_{t-i}'\right] (\Phi_1')'$$

$$= \sum_{i=0}^{\infty} \Phi_1^i \Sigma (\Phi_1')'$$

$$\mathrm{vec}\left(\mathrm{E}\left[(\mathbf{Y}_t - \mu)(\mathbf{Y}_t - \mu)'\right]\right) = (\mathbf{I}_{k^2} - \Phi_1 \otimes \Phi_1)^{-1} \mathrm{vec}(\Sigma)$$

where $\mu = (\mathbf{I}_k - \Phi_1)^{-1} \Phi_0$. The similarity between the long-run covariance of a VAR(1) and the long-run variance of a univariate autoregression, $\sigma^2/(1 - \phi_1^2)$, are less pronounced. The difference between

these expressions arises since matrix multiplication is non-commutative ($\mathbf{AB} \neq \mathbf{BA}$, in general). The final line makes use of the *vec* (vector) operator to compactly express the long-run covariance. The vec operator and a Kronecker product stack the elements of a matrix product into a single column.[3] The eigenvalues of Φ_1 also affect the long-run covariance, and if any are close to 1, the long-run covariance is large since the maximum eigenvalue determines the persistence of shocks. All things equal, more persistence lead to larger long-run covariances since the effect of any shock last longer.

5.1.2.4 Autocovariance

The autocovariances of a vector-valued stochastic process are defined

Definition 5.7 (Autocovariance). The autocovariance matrices of k by 1 vector-valued covariance stationary stochastic process $\{\mathbf{Y}_t\}$ are defined

$$\Gamma_s = \mathrm{E}[(\mathbf{Y}_t - \mu)(\mathbf{Y}_{t-s} - \mu)'] \tag{5.10}$$

and

$$\Gamma_{-s} = \mathrm{E}[(\mathbf{Y}_t - \mu)(\mathbf{Y}_{t+s} - \mu)'] \tag{5.11}$$

where $\mu = \mathrm{E}[\mathbf{Y}_t] = \mathrm{E}[\mathbf{Y}_{t-j}] = \mathrm{E}[\mathbf{Y}_{t+j}]$.

The structure of the autocovariance function is the first significant deviation from the univariate time-series analysis in chapter 4. Vector autocovarianes are reflected, and so are symmetric only when transposed. Specifically,

---

[3]The vec of a matrix A is defined:

Definition 5.5 (vec). Let $\mathbf{A} = [a_{ij}]$ be an m by n matrix. The vec operator (also known as the *stack* operator) is defined

$$\mathrm{vec}\,\mathbf{A} = \begin{bmatrix} \mathbf{a}_1 \\ \mathbf{a}_2 \\ \vdots \\ \mathbf{a}_n \end{bmatrix} \tag{5.9}$$

where $\mathbf{a}_j$ is the j$^{\text{th}}$ column of the matrix $\mathbf{A}$.

The Kronecker Product is defined:

Definition 5.6 (Kronecker Product). Let $\mathbf{A} = [a_{ij}]$ be an m by n matrix, and let $\mathbf{B} = [b_{ij}]$ be a k by l matrix. The Kronecker product is defined

$$\mathbf{A} \otimes \mathbf{B} = \begin{bmatrix} a_{11}\mathbf{B} & a_{12}\mathbf{B} & \dots & a_{1n}\mathbf{B} \\ a_{21}\mathbf{B} & a_{22}\mathbf{B} & \dots & a_{2n}\mathbf{B} \\ \vdots & \vdots & \vdots & \vdots \\ a_{m1}\mathbf{B} & a_{m2}\mathbf{B} & \dots & a_{mn}\mathbf{B} \end{bmatrix}$$

and has dimension mk by nl.

It can be shown that

Theorem 5.4 (Kronecker and vec of a product). *Let* $\mathbf{A}$, $\mathbf{B}$ *and* $\mathbf{C}$ *be conformable matrices as needed. Then*

$$\mathrm{vec}\,(\mathbf{ABC}) = (\mathbf{C}' \otimes \mathbf{A})\,\mathrm{vec}\,\mathbf{B}$$

$$\Gamma_s \neq \Gamma_{-s}$$

but[4]

$$\Gamma_s = \Gamma'_{-s}.$$

In contrast, the autocovariances of stationary scalar processes satisfy $\gamma_s = \gamma_{-s}$. Computing the autocovariances uses the backward substitution form so that

$$\Gamma_s = \mathrm{E}\left[(\mathbf{Y}_t - \mu)(\mathbf{Y}_{t-s} - \mu)'\right] = \mathrm{E}\left[\left(\sum_{i=0}^{\infty} \Phi_1^i \varepsilon_{t-i}\right)\left(\sum_{i=0}^{\infty} \Phi_1^i \varepsilon_{t-s-i}\right)'\right] \tag{5.12}$$

$$= \mathrm{E}\left[\left(\sum_{i=0}^{s-1} \Phi_1^i \varepsilon_{t-i}\right)\left(\sum_{i=0}^{\infty} \Phi_1^i \varepsilon_{t-s-i}\right)'\right]$$

$$+ \mathrm{E}\left[\left(\sum_{i=0}^{\infty} \Phi_1^s \Phi_1^i \varepsilon_{t-s-i}\right)\left(\sum_{i=0}^{\infty} \Phi_1^i \varepsilon_{t-s-i}\right)'\right] \tag{5.13}$$

$$= \mathbf{0} + \Phi_1^s \mathrm{E}\left[\left(\sum_{i=0}^{\infty} \Phi_1^i \varepsilon_{t-s-i}\right)\left(\sum_{i=0}^{\infty} \Phi_1^i \varepsilon_{t-s-i}\right)'\right]$$

$$= \Phi_1^s \mathrm{V}\left[\mathbf{Y}_t\right]$$

and

$$\Gamma_{-s} = \mathrm{E}\left[(\mathbf{Y}_t - \mu)(\mathbf{Y}_{t+s} - \mu)'\right] = \mathrm{E}\left[\left(\sum_{i=0}^{\infty} \Phi_1^i \varepsilon_{t-i}\right)\left(\sum_{i=0}^{\infty} \Phi_1^i \varepsilon_{t+s-i}\right)'\right] \tag{5.14}$$

$$= \mathrm{E}\left[\left(\sum_{i=0}^{\infty} \Phi_1^i \varepsilon_{t-i}\right)\left(\sum_{i=0}^{\infty} \Phi_1^s \Phi_1^i \varepsilon_{t-i}\right)'\right]$$

$$+ \mathrm{E}\left[\left(\sum_{i=0}^{\infty} \Phi_1^i \varepsilon_{t-i}\right)\left(\sum_{i=0}^{s-1} \Phi_1^i \varepsilon_{t+s-i}\right)'\right] \tag{5.15}$$

$$= \mathrm{E}\left[\left(\sum_{i=0}^{\infty} \Phi_1^i \varepsilon_{t-i}\right)\left(\sum_{i=0}^{\infty} \varepsilon'_{t-i} \left(\Phi_1'\right)^i \left(\Phi_1'\right)^s\right)\right] + \mathbf{0}$$

$$= \mathrm{E}\left[\left(\sum_{i=0}^{\infty} \Phi_1^i \varepsilon_{t-i}\right)\left(\sum_{i=0}^{\infty} \varepsilon'_{t-i} \left(\Phi_1'\right)^i\right)\right] \left(\Phi_1'\right)^s$$

$$= \mathrm{V}\left[\mathbf{Y}_t\right] \left(\Phi'_1\right)^s$$

where $\mathrm{V}[\mathbf{Y}_t]$ is the symmetric covariance matrix of the VAR. Like most properties of a VAR, the autocovariance function of a VAR(1) closely resembles that of an AR(1): $\gamma_s = \phi_1^{|s|} \sigma^2 / (1 - \phi_1^2) = \phi_1^{|s|} \mathrm{V}[Y_t]$.

---

[4]This follows directly from the property of a transpose that if **A** and **B** are compatible matrices, $(\mathbf{AB})' = \mathbf{B}'\mathbf{A}'$.

5.2   Companion Form

Any stationary VAR(P) can be rewritten as a VAR(1). Suppose $\{\mathbf{Y}_t\}$ follows a VAR(P) process,

$$\mathbf{Y}_t = \Phi_0 + \Phi_1 \mathbf{Y}_{t-1} + \Phi_2 \mathbf{Y}_{t-2} + \ldots + \Phi_P \mathbf{Y}_{t-P} + \varepsilon_t.$$

By subtracting the mean and stacking P lags of $\mathbf{Y}_t$ into a large column vector denoted $\mathbf{Z}_t$, a VAR(P) is equivalently expressed as a VAR(1) using the companion form.

Definition 5.8 (Companion Form of a VAR(P)). Let $\mathbf{Y}_t$ follow a VAR(P) given by

$$\mathbf{Y}_t = \Phi_0 + \Phi_1 \mathbf{Y}_{t-1} + \Phi_2 \mathbf{Y}_{t-2} + \ldots + \Phi_P \mathbf{Y}_{t-P} + \varepsilon_t$$

where ε_t is a vector white noise process and $\mu = \left(\mathbf{I} - \sum_{p=1}^{P} \Phi_p\right)^{-1} \Phi_0 = \mathrm{E}[\mathbf{Y}_t]$ is finite. The companion form is

$$\mathbf{Z}_t = \Upsilon \mathbf{Z}_{t-1} + \xi_t \tag{5.16}$$

where

$$\mathbf{Z}_t = \begin{bmatrix} \mathbf{Y}_t - \mu \\ \mathbf{Y}_{t-1} - \mu \\ \vdots \\ \mathbf{Y}_{t-P+1} - \mu \end{bmatrix}, \tag{5.17}$$

$$\Upsilon = \begin{bmatrix} \Phi_1 & \Phi_2 & \Phi_3 & \ldots & \Phi_{P-1} & \Phi_P \\ \mathbf{I}_k & \mathbf{0} & \mathbf{0} & \ldots & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_k & \mathbf{0} & \ldots & \mathbf{0} & \mathbf{0} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \ldots & \mathbf{I}_k & \mathbf{0} \end{bmatrix} \tag{5.18}$$

and

$$\xi_t = \begin{bmatrix} \varepsilon_t \\ \mathbf{0} \\ \vdots \\ \mathbf{0} \end{bmatrix}, \mathrm{E}[\xi_t \xi_t'] = \begin{bmatrix} \Sigma & \mathbf{0} & \ldots & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \ldots & \mathbf{0} \\ \vdots & \vdots & \vdots & \vdots \\ \mathbf{0} & \mathbf{0} & \ldots & \mathbf{0} \end{bmatrix}.. \tag{5.19}$$

The properties of a VAR(P) are identical to that of its companion form VAR(1). For example, VAR(P) is covariance stationary if all of the eigenvalues of Υ - there are $k \times P$ of them - are less than one in absolute value (modulus if complex).[5]

5.3   Empirical Examples

Two examples from the macrofinance literature are used throughout this chapter to illustrate the application of VARs.

---

[5]The companion form is also useful when working with univariate AR(P) models. An AR(P) can be equivalently expressed as a VAR(1), which simplifies computing properties such as the long-run variance and autocovariances.

5.3.1   Example: The interaction of stock and bond returns

Stocks and bonds are often thought to hedge one another. VARs provide a simple method to determine whether their returns are linked through time. Consider the VAR(1)

$$\begin{bmatrix} VWM_t \\ TERM_t \end{bmatrix} = \begin{bmatrix} \phi_{01} \\ \phi_{02} \end{bmatrix} + \begin{bmatrix} \phi_{11,1} & \phi_{12,1} \\ \phi_{21,1} & \phi_{22,1} \end{bmatrix} \begin{bmatrix} VWM_{t-1} \\ TERM_{t-1} \end{bmatrix} + \begin{bmatrix} \varepsilon_{1,t} \\ \varepsilon_{2,t} \end{bmatrix}$$

where VWM_t is the return on the value-weighted-market portfolio and $TERM_t$ is the return on a portfolio that is long the 10-year and short the 1-year U.S. government bond. The VAR contains a model for stock returns

$$VWM_t = \phi_{01} + \phi_{11,1}VWM_{t-1} + \phi_{12,1}TERM_{t-1} + \varepsilon_{1,t}$$

and a model for the return on the term premium,

$$TERM_t = \phi_{01} + \phi_{21,1}VWM_{t-1} + \phi_{22,1}TERM_{t-1} + \varepsilon_{2,t}.$$

Since these models do not share any parameters, the coefficient can be estimated equation-by-equation using OLS.[6] A VAR(1) is estimated using monthly return data (multiplied by 12) for the *VWM* from CRSP and the 10-year constant maturity treasury yield from FRED covering the period February 1962 until December 2018.[7]

$$\begin{bmatrix} VWM_t \\ TERM_t \end{bmatrix} = \begin{bmatrix} \underset{(0.000)}{0.801} \\ \underset{(0.041)}{0.232} \end{bmatrix} + \begin{bmatrix} \underset{(0.122)}{0.059} & \underset{(0.004)}{0.166} \\ \underset{(0.000)}{-0.104} & \underset{(0.002)}{0.116} \end{bmatrix} \begin{bmatrix} VWM_{t-1} \\ TERM_{t-1} \end{bmatrix} + \begin{bmatrix} \varepsilon_{1,t} \\ \varepsilon_{2,t} \end{bmatrix}$$

The p-value of each coefficient is reported in parenthesis. The estimates indicate that stock returns are not predictable using past stock returns but are predictable using the returns on the lagged term premium: positive returns on the term premium lead increase expected returns in stocks. In contrast, positive returns in equities decrease the expected return on the term premium. The annualized long-run mean can be computed from the estimated parameters as

$$12 \times \left(\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} - \begin{bmatrix} 0.059 & 0.166 \\ -0.104 & 0.116 \end{bmatrix} \right)^{-1} \begin{bmatrix} 0.801 \\ 0.232 \end{bmatrix} = \begin{bmatrix} 10.558 \\ 1.907 \end{bmatrix}.$$

These model-based estimates are similar to the sample averages of returns of 10.57 and 1.89 for *VWM* and *TERM*, respectively.

5.3.2   Example: Monetary Policy VAR

VARs are widely used in macrofinance to model closely related macroeconomic variables. This example uses a 3-variable VAR containing the unemployment rate, the effective federal funds rate, which is the rate that banks use to lend to each other, and inflation. Inflation is measured using the implicit

---

[6]Theoretical motivations often lead to cross-parameter equality restrictions in VARs. These models cannot be estimated equation-by-equation. A VAR subject to linear equality restrictions can be estimated using a system OLS estimator.

[7]The yields of the bonds are converted to prices, and then returns are computed as the log difference of the prices plus accrued interest.

| | Raw Data | | |
	$\Delta \ln \text{UNEMP}_{t-1}$	FF_{t-1}	ΔINF_{t-1}
$\Delta \ln \text{UNEMP}_t$	0.624 (0.000)	0.015 (0.001)	0.016 (0.267)
FF_t	−0.816 (0.000)	0.979 (0.000)	−0.045 (0.317)
ΔINF_t	−0.501 (0.010)	−0.009 (0.626)	−0.401 (0.000)

| | Standardized Series | | |
	$\Delta \ln \text{UNEMP}_{t-1}$	FF_{t-1}	ΔINF_{t-1}
$\Delta \ln \text{UNEMP}_t$	0.624 (0.000)	0.153 (0.001)	0.053 (0.267)
FF_t	−0.080 (0.000)	0.979 (0.000)	−0.015 (0.317)
ΔINF_t	−0.151 (0.010)	−0.028 (0.626)	−0.401 (0.000)

Table 5.1: Parameter estimates from the monetary policy VAR. The top panel contains estimates using original, unmodified values while the bottom panel contains estimates from data standardized to have unit variance. While the magnitudes of many coefficients change, the p-values and the eigenvalues of the parameter matrices are identical, and the parameters are roughly comparable since the series have the same variance.

GDP price deflator. Two of the three variables, the unemployment and inflation rates, appear to be nonstationary when tested using an ADF test, and so are differenced.[8]

Using a VAR(1) specification, the model can be described

$$\begin{bmatrix} \Delta \text{UNEMP}_t \\ \text{FF}_t \\ \Delta \text{INF}_t \end{bmatrix} = \Phi_0 + \Phi_1 \begin{bmatrix} \Delta \text{UNEMP}_{t-1} \\ \text{FF}_{t-1} \\ \Delta \text{INF}_{t-1} \end{bmatrix} + \begin{bmatrix} \varepsilon_{1,t} \\ \varepsilon_{2,t} \\ \varepsilon_{3,t} \end{bmatrix}.$$

Two sets of parameters are presented in Table 5.1. The top panel contains estimates using non-scaled data. The bottom panel contains estimates from data where each series is standardized to have unit variance. Standardization produces coefficients that have comparable magnitudes. Despite this transformation and very different parameter estimates, the p-values remain unchanged since OLS t-stats are invariant to rescalings of this type. The eigenvalues of the two parameter matrices are identical, and so the estimate of the persistence of the process is not affected by standardizing the data.

---

[8]All three series, UNRATE (unemployment), DFF (Federal Funds), and GDPDEF (deflator), are available in FRED. The unemployment and Federal Funds rates are aggregated to quarterly by taking the mean of all observations within a quarter. The inflation rate is computed from the deflator as $400 \ln\left(\text{GDPDEF}_t / \text{GDPDEF}_{t-1}\right)$.

5.4   VAR forecasting

Constructing forecasts of a vector time series is identical to constructing the forecast from a single time series. h-step forecasts are recursively constructed starting with $E_t[\mathbf{Y}_{t+1}]$, using $E_t[\mathbf{Y}_{t+1}]$ to construct $E_t[\mathbf{Y}_{t+2}]$, and continuing until $E_t[\mathbf{Y}_{t+h}]$.

Recall that the h-step ahead forecast, $\hat{Y}_{t+h|t}$ in an AR(1) is

$$E_t[Y_{t+h}] = \sum_{j=0}^{h-1} \phi_1^j \phi_0 + \phi_1^h Y_t.$$

The h-step ahead forecast of a VAR(1) , $\hat{\mathbf{Y}}_{t+h|t}$, has the same structure, and is

$$E_t[\mathbf{Y}_{t+h}] = \sum_{j=0}^{h-1} \Phi_1^j \Phi_0 + \Phi_1^h \mathbf{Y}_t.$$

This formula can be used to produce multi-step forecast of any VAR using the companion form.

In practice, it is simpler to compute the forecasts using the deviations form of the VAR since it includes no intercept,

$$\tilde{\mathbf{Y}}_t = \Phi_1 \tilde{\mathbf{Y}}_{t-1} + \Phi_2 \tilde{\mathbf{Y}}_{t-2} + \ldots + \Phi_P \tilde{\mathbf{Y}}_{t-P} + \varepsilon_t,$$

where $\mu = (\mathbf{I}_k - \Phi_1 - \ldots - \Phi_P)^{-1} \Phi_0$ and $\tilde{\mathbf{Y}}_t = \mathbf{Y}_t - \mu$ are mean 0. The h-step forecasts from the deviations form are computed using the recurrence

$$E_t[\tilde{\mathbf{Y}}_{t+h}] = \Phi_1 E_t[\tilde{\mathbf{Y}}_{t+h-1}] + \Phi_2 E_t[\tilde{\mathbf{Y}}_{t+h-2}] + \ldots + \Phi_P E_t[\tilde{\mathbf{Y}}_{t+h-P}].$$

starting at $E_t[\tilde{\mathbf{Y}}_{t+1}]$. Using the forecast of $E_t[\tilde{\mathbf{Y}}_{t+h}]$, the h-step ahead forecast of $\mathbf{Y}_{t+h}$ is constructed by adding the long-run mean, $E_t[\mathbf{Y}_{t+h}] = \mu + E_t[\tilde{\mathbf{Y}}_{t+h}]$.

5.4.1   Example: Monetary Policy VAR

Forecasts from VARs incorporate information beyond the history of a single time series. Table 5.2 contains the relative Mean Square Error of out-of-sample forecasts for the three variables in the policy VAR. Each set of forecasts is produced by recursively estimating model parameters using a minimum of 50% of the available sample. Forecasts are produced for up to 8 quarters ahead. Each series is also forecast using a univariate AR model.

The out-of-sample MSE of the forecasts from a model is defined

$$\text{MSE} = 1/T - h - R \sum_{t=R}^{T-h} \left(Y_{t+h} - \hat{Y}_{t+h|t}\right)^2$$

where R is the size of the initial in-sample period, Y_{t+h} is the realization of the variable in period $t+h$, and $\hat{Y}_{t+h|t}$ is the h-step ahead forecast produced at time t. The relative MSE is defined as

$$\text{Relative MSE} = \frac{\text{MSE}}{\text{MSE}_{bm}}$$

Horizon	Series	VAR		AR	
		Restricted	Unrestricted	Restricted	Unrestricted
1	Unemployment	0.522	**0.520**	0.520	0.521
	Fed. Funds Rate	**0.887**	0.903	0.917	0.927
	Inflation	0.869	0.868	**0.839**	0.839
2	Unemployment	0.716	**0.710**	0.743	0.745
	Fed. Funds Rate	**0.923**	0.943	*1.102*	*1.119*
	Inflation	*1.082*	*1.081*	*1.030*	*1.030*
4	Unemployment	0.872	**0.861**	0.962	0.965
	Fed. Funds Rate	**0.952**	0.976	*1.071*	*1.098*
	Inflation	*1.000*	0.999	0.998	**0.998**
8	Unemployment	0.820	**0.806**	0.971	0.977
	Fed. Funds Rate	**0.974**	*1.007*	*1.058*	*1.105*
	Inflation	*1.001*	1.000	0.998	**0.997**

Table 5.2: Relative out-of-sample Mean Square Error for forecasts between 1 and 8-quarters ahead. The benchmark model is a constant for the unemployment rate and the inflation rate and a random walk for the Federal Funds rate. Model parameters are recursively estimated, and forecasts are produced once 50% of the available sample. Model order is selected using the BIC.

where MSE_{bm} is the out-of-sample MSE of a benchmark model. The Federal Funds rate is modeled in levels, and so the benchmark model is a random walk. The other two series are differenced, and so use the historical mean (an AR(0)) as the benchmark model. The number of lags in either the VAR or the AR is selected by minimizing the BIC (see Section 5.5).

Each model is estimated using two methods, the standard estimator and a restricted estimator where the long-run mean forced to match the in-sample mean. The restricted model is estimated in two steps. First, the sample mean is subtracted, and then the model is estimated without a constant. The forecasts are then constructed using the sample mean plus the forecast of the demeaned data. The two-step estimator ensures that the model mean reverts to the historical average. The unrestricted model jointly estimates the intercept with the parameters that capture the dynamics and so does not revert (exactly) to the sample mean even over long horizons. These two method can produce qualitatively different forecasts in persistent time series due to differences in the average values of the data used as lags ($\bar{\mathbf{Y}}_{t-j} = (T-P)^{-1}\sum_{t=P-j+1}^{T-j}\mathbf{Y}_t$ for $j = 1, 2, \ldots, P$) and the average value of the contemporaneous values ($\bar{\mathbf{Y}}_t = (T-P)^{-1}\sum_{t=P+1}^{T}\mathbf{Y}_t$). The two-step estimator uses the same mean value for both, $\bar{\mathbf{Y}} = T^{-1}\sum_{t=1}^{P}\mathbf{Y}_t$.

The VAR performs well in this forecasting problem. It produced the lowest MSE in 7 of 12 horizon-series combinations. When it is not the best model, it performs only slightly worse than the autoregression. Ultimately, the choice of a model to use in forecasting applications – either multivariate or univariate – is an empirical question that is best answered using in-sample analysis and pseudo-out-of-sample forecasting.

5.5   Estimation and Identification

Understanding the dependence structure in VAR models requires additional measures of cross-variable relationships. The cross-correlation function (CCF) and partial cross-correlation function (PCCF) extend the autocorrelation and partial autocorrelation functions used to identify the model order in a single time series.

Definition 5.9 (Cross-correlation)**.** The sth cross correlations between two covariance stationary series $\{x_t\}$ and $\{Y_t\}$ are defined

$$\rho_{xy,s} = \frac{E[(x_t - \mu_x)(Y_{t-s} - \mu_y)]}{\sqrt{V[x_t]V[Y_t]}} \tag{5.20}$$

and

$$\rho_{yx,s} = \frac{E[(Y_t - \mu_y)(x_{t-s} - \mu_x)]}{\sqrt{V[x_t]V[Y_t]}} \tag{5.21}$$

where the order of the indices indicates the lagged variable, $E[Y_t] = \mu_y$ and $E[x_t] = \mu_x$.

Cross-correlations, unlike autocorelations, are not symmetric in the order of the arguments. Partial cross-correlations are defined using a similar extension of partial autocorrelation as the correlation between x_t and Y_{t-s} controlling for $Y_{t-1}, \ldots, Y_{t-(s-1)}$.

Definition 5.10 (Partial Cross-correlation)**.** The partial cross-correlations between two covariance stationary series $\{x_t\}$ and $\{Y_t\}$ are defined as the population values of the coefficients $\varphi_{xy,s}$ in the regression

$$x_t = \phi_0 + \phi_{x1}x_{t-1} + \ldots + \phi_{xs-1}x_{t-(s-1)} + \phi_{y1}Y_{t-1} + \ldots + \phi_{ys-1}Y_{t-(s-1)} + \varphi_{xy,s}Y_{t-s} + \varepsilon_{x,t} \tag{5.22}$$

and $\varphi_{yx,s}$ in the regression

$$Y_t = \phi_0 + \phi_{y1}Y_{t-1} + \ldots + \phi_{ys-1}Y_{t-(s-1)} + \phi_{x1}x_{t-1} + \ldots + \phi_{xs-1}x_{t-(s-1)} + \varphi_{yx,s}x_{t-s} + \varepsilon_{y,t} \tag{5.23}$$

where the order of the indices indicates which lagged variable. In a k-variable VAR, the PCCF of series i with respect to series j is the population value of the coefficient $\varphi_{Y_iY_js}$ in the regression

$$Y_{it} = \phi_0 + \phi_1' Y_{t-1} + \ldots + \phi_{s-1}' + \varphi_{Y_iY_js}Y_{jt-s} + \varepsilon_i$$

where ϕ_j are 1 by k vectors of parameters.

The controls in the sth partial cross-correlation are included variables in a VAR(s-1). If the data are generated by a VAR(P), then the sth partial cross-correlation is 0 whenever $s > P$. This behavior is analogous to the behavior of the PACF in an AR(P) model. The PCCF is a useful diagnostic to identify the order of a VAR and for verifying the order of estimated models when applied to residuals.

Figure 5.1 plots 1,000 simulated data points from a high-order bivariate VAR. One component of the VAR follows a HAR(22) process with no spillovers from the other component. The second component is substantially driven by both spillovers from the HAR and its own innovation. The complete specification of the VAR(22) is
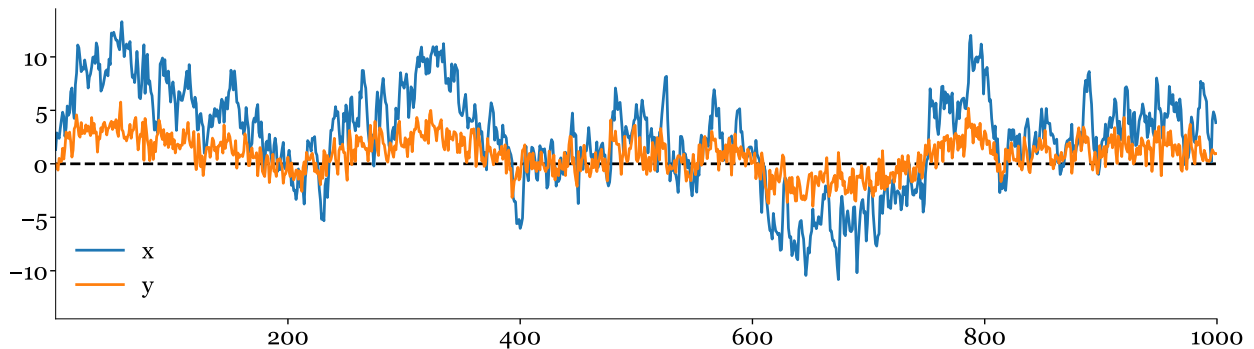
Simulated data from a VAR(22)

Figure 5.1: Simulated data from the VAR(22) in eq. (5.24). Both processes are stationary but highly persistent and have a high degree of comovement.

$$\begin{bmatrix} x_t \\ Y_t \end{bmatrix} = \begin{bmatrix} 0.5 & 0.9 \\ .0 & 0.47 \end{bmatrix} \begin{bmatrix} x_{t-1} \\ Y_{t-1} \end{bmatrix} + \sum_{i=2}^{5} \begin{bmatrix} 0 & 0 \\ 0 & 0.06 \end{bmatrix} \begin{bmatrix} x_{t-i} \\ Y_{t-i} \end{bmatrix} + \sum_{j=6}^{22} \begin{bmatrix} 0 & 0 \\ 0 & 0.01 \end{bmatrix} \begin{bmatrix} x_{t-j} \\ Y_{t-j} \end{bmatrix} + \begin{bmatrix} \varepsilon_{x,t} \\ \varepsilon_{y,t} \end{bmatrix}.$$

(5.24)

Figure 5.2 contains plots of the theoretical ACF and CCF (cross-correlation function) of this VAR. Both ACFs and CCFs indicate that the series are highly persistent. They also show that both variables are a strong predictor of either at any lag since the squared correlation can be directly interpretable as an R^2. Figure 5.3 contains plots of the partial auto- and cross-correlation function. These are markedly different from the ACFs and CCFs. The PACF and PCCF of x both cut off after one lag. This happens since x has 0 coefficients on all lagged values after the first. The PACF and PCCF of y are more complex. The PACF resembles the step-function of the coefficients in the HAR model. It cuts off sharply after 22 lags since this is the order of the VAR. The PCCF of y is also non-zero for many lags, and only cuts off after 21. The reduction in the cut-off is due to the structure of the VAR where x is only exposed to the lagged value of y at the first lag, and so the dependence is reduced by one.

These new definitions enable the key ideas of the Box-Jenkins methodology to be extended to vector processes. While this extension is technically possible, using the ACF, PACF, CCF, and PCCF to determine the model lag length is difficult. The challenge of graphical identification of the order is especially daunting in specifications with more than two variables since there are many dependence measures to inspect – a k-dimensional stochastic process has $2(k^2 - k)$ distinct auto- and cross-correlation functions.

The standard approach is to adopt the approach advocated in Sims (1980). The VAR specification should include all variables that theory indicates are relevant, and the lag length should be chosen so that the model has a high likelihood of capturing all of the dynamics. Once the maximum value of the lag length is chosen, a general-to-specific search can be conducted to reduce the model order, or an information criterion can be used to select an appropriate lag length. In a VAR, the Akaike IC, Hannan and Quinn (1979) IC and the Schwarz/Bayesian IC are
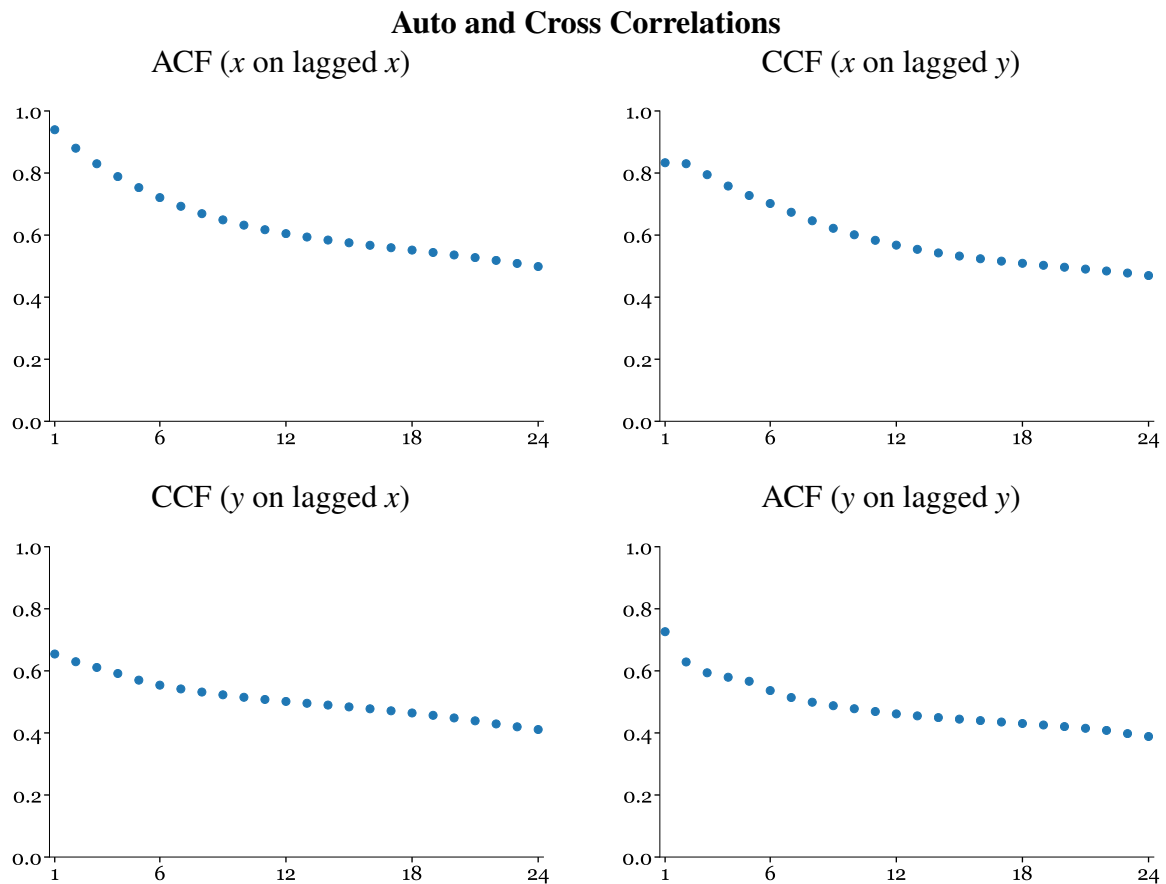
Auto and Cross Correlations

Figure 5.2: The four panels contain the ACFs and CCFs of the VAR(22) process in eq. (5.24).

$$\text{AIC:} \qquad \ln|\hat{\Sigma}(P)| + k^2 P \frac{2}{T}$$

$$\text{HQIC:} \qquad \ln|\hat{\Sigma}(P)| + k^2 P \frac{2\ln\ln T}{T}$$

$$\text{BIC:} \qquad \ln|\hat{\Sigma}(P)| + k^2 P \frac{\ln T}{T}$$

where $\hat{\Sigma}(P)$ is the covariance of the residuals estimated using a VAR(P) and $|\cdot|$ is the determinant.[9] All models must use the same values on the left-hand-side irrespective of the lags included when choosing the lag length. In practice, it is necessary to adjust the sample when estimating the parameters of models with fewer lags than the maximum allowed. For example, when comparing models with up to 2 lags, the largest model is estimated by fitting observations $3,4,\ldots,T$ since two lags are lost when constructing the right-hand-side variables. The 1-lag model should also fit observations $3,4,\ldots,T$ and so observation 1 is excluded from the model since it is not needed as a lagged variable.

---

[9] $\ln|\hat{\Sigma}|$ is, up to an additive constant, the Gaussian log-likelihood divided by T. These three information criteria are all special cases of the usual information criteria for log-likelihood models which take the form $-L + P_{IC}$ where P_{IC} is the penalty which depends on the number of estimated parameters in the model and the information criterion.

Partial Auto and Cross Correlations

PACF (*x* on lagged *x*)   PCCF (*x* on lagged *y*)

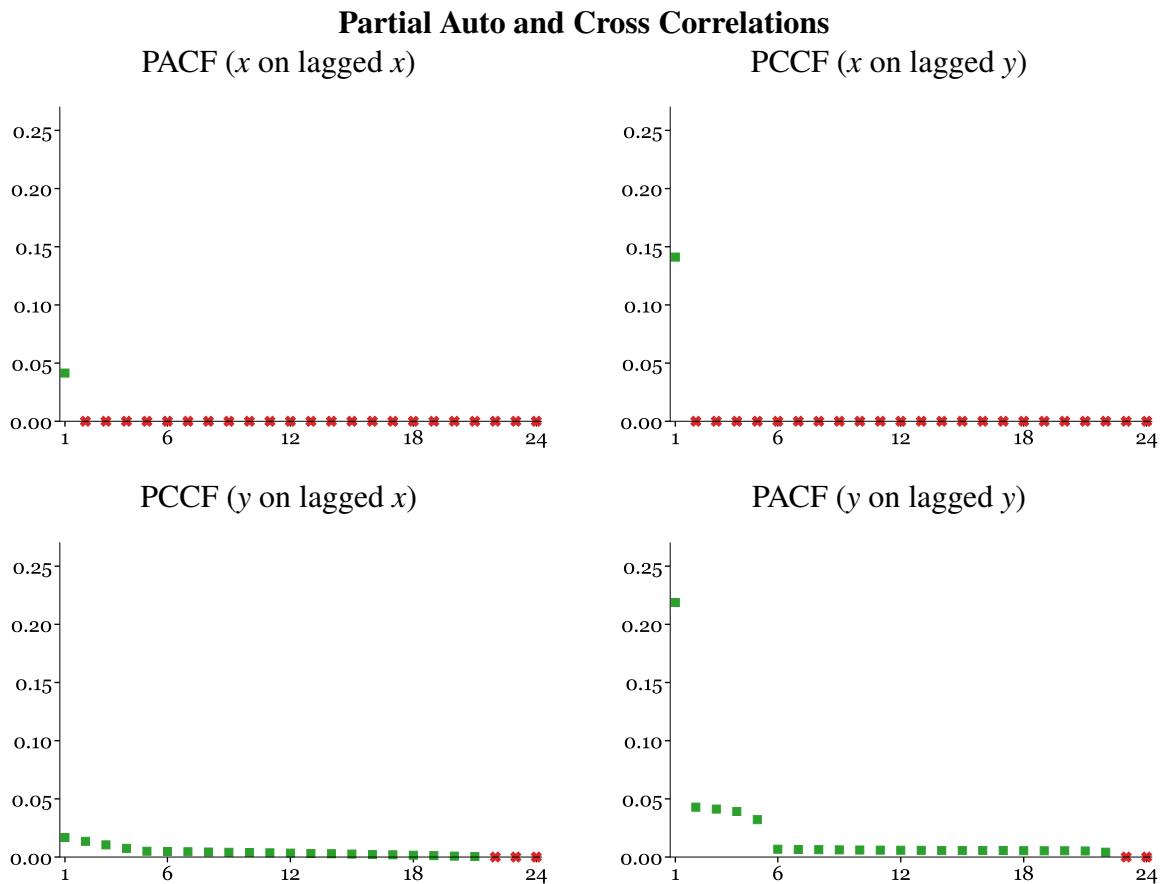PCCF (*y* on lagged *x*)   PACF (*y* on lagged *y*)

Figure 5.3: The four panels contain the PACFs and PCCFs of the VAR(22) process in eq. (5.24). Values marked with a red **x** are exactly 0.

The lag length should be chosen to minimize one of the criteria. The BIC has the most substantial penalty term and so always chooses a (weakly) smaller model than the HQIC. The AIC has the smallest penalty, and so selects the largest model of the three ICs. Ivanov and Kilian (2005) recommend the AIC for monthly models and the HQIC for quarterly models unless the sample size is less than 120 quarters. In short samples, the BIC is preferred. Their recommendation is based on the accuracy of the impulse response function, and so may not be ideal in other applications, e.g., forecasting.

Alternatively, a likelihood ratio test can be used to test whether to specifications are equivalent. The LR test statistic is

$$(T - P_2 k^2) \left(\ln|\hat{\Sigma}(P_1)| - \ln|\hat{\Sigma}(P_2)| \right) \overset{A}{\sim} \chi^2_{(P_2 - P_1)k^2},$$

where P_1 is the number of lags in the restricted (smaller) model, P_2 is the number of lags in the unrestricted (larger) model and k is the dimension of $\mathbf{Y}_t$. Since model 1 is a restricted version of model 2, its covariance is larger and so this statistic is always positive. The $-P_2 k^2$ term in the log-likelihood is a degree of freedom correction that generally improves small-sample performance of the test. Ivanov and Kilian (2005) recommend against using sequential likelihood ratio testing in lag length selection.

Lag Length	AIC	HQIC	BIC	LR	P-val
0	4.014	3.762	3.605	925	0.000
1	0.279	0.079	0.000$^{\blacktriangledown\blacktriangle}$	39.6	0.000
2	0.190	0.042	0.041	40.9	0.000
3	0.096	0.000$^{\blacktriangledown}$	0.076	29.0	0.001
4	0.050$^{\blacktriangledown}$	0.007	0.160	7.34	0.602$^{\blacktriangledown}$
5	0.094	0.103	0.333	29.5	0.001
6	0.047	0.108	0.415	13.2	0.155
7	0.067	0.180	0.564	32.4	0.000
8	0.007	0.172$^{\blacktriangle}$	0.634	19.8	0.019
9	0.000$^{\blacktriangle}$	0.217	0.756	7.68	0.566$^{\blacktriangle}$
10	0.042	0.312	0.928	13.5	0.141
11	0.061	0.382	1.076	13.5	0.141
12	0.079	0.453	1.224	–	–

Table 5.3: Normalized values for the AIC, HQIC, and BIC in a Monetary Policy VAR. The information criteria are normalized by subtracting the smallest value from each column. The LR and P-value in each row are for a test with the null that the coefficient on lag $l+1$ are all zero ($H_0 : \Phi_{l+1} = \mathbf{0}$) and the alternative $H_1 : \Phi_{l+1} \neq \mathbf{0}$. Values marked with $\blacktriangledown$ indicate the lag length selected using a specific-to-general search. Values marked with $\blacktriangle$ indicate the lag length selected using general-to-specific.

5.5.1   Example: Monetary Policy VAR

The Monetary Policy VAR is used to illustrate lag length selection. The information criteria, log-likelihoods, and p-values from the LR tests are presented in Table 5.3. This table contains the AIC, HQIC, and BIC values for lags 0 (no dynamics) through 12 as well as likelihood ratio test results for testing l lags against $l+1$. Note that the LR and P-value corresponding to lag l test the null that the fit using l lags is equivalent to the fit using $l+1$ lags. Using the AIC, 9 lags produces the smallest value and is selected in a general-to-specific search. A specific-to-general search stops at 4 lags since the AIC of 5 lags is larger than the AIC of 4. The HQIC chooses 3 lags in a specific-to-general search and 9 in a general-to-specific search. The BIC selects a single lag irrespective of the search direction. A general-to-specific search using the likelihood ratio chooses 9 lags, and a hypothesis-test-based specific-to-general procedure chooses 4. The specific-to-general stops at 4 lags since the null $H_0 : P = 4$ tested against the alternative that $H_1 : P = 5$ has a p-value of 0.602, which indicates that these models provide similar fits of the data.

Finally, the information criteria are applied in a "global search" that evaluates models using every combination of lags up to 12. This procedure fits a total of 4096 VARs (which only requires a few seconds on a modern computer), and the AIC, HQIC, and the BIC are computed for each.[10] Using this methodology, the AIC search selected lags 1–3 and 7–9, the HQIC selects lags 1–3, 6, and 8, and the BIC continues to select a parsimonious model that includes only the first lag. Search procedures of this type are computationally viable for checking up to 20 lags.

---

[10]For a maximum lag length of L, 2^L models must be estimated.

5.6 Granger causality

Granger causality (GC, also known as *prima facia* causality) is the first concept exclusive to vector analysis. GC is the standard method to determine whether one variable is useful in predicting another and evidence of Granger causality it is a good indicator that a VAR, rather than a univariate model, is needed.

Granger causality is defined in the negative.

Definition 5.11 (Granger causality). A scalar random variable x_t does not Granger cause Y_t if $\mathrm{E}[Y_t|x_{t-1},Y_{t-1},x_{t-2},$ $\mathrm{E}[Y_t|,Y_{t-1},Y_{t-2},\ldots].$[11] That is, x_t does not Granger cause Y_t if the forecast of Y_t is the same whether conditioned on past values of x_t or not.

Granger causality can be simply illustrated in a bivariate VAR.

$$\begin{bmatrix} x_t \\ Y_t \end{bmatrix} = \begin{bmatrix} \phi_{11,1} & \phi_{12,1} \\ \phi_{21,1} & \phi_{22,1} \end{bmatrix} \begin{bmatrix} x_{t-1} \\ Y_{t-1} \end{bmatrix} + \begin{bmatrix} \phi_{11,2} & \phi_{12,2} \\ \phi_{21,2} & \phi_{22,2} \end{bmatrix} \begin{bmatrix} x_{t-2} \\ Y_{t-2} \end{bmatrix} + \begin{bmatrix} \varepsilon_{1,t} \\ \varepsilon_{2,t} \end{bmatrix}$$

In this model, if $\phi_{21,1} = \phi_{21,2} = 0$ then x_t does not Granger cause Y_t. Note that x_t not Granger causing Y_t says nothing about whether Y_t Granger causes x_t.

An important limitation of GC is that it does not account for indirect effects. For example, suppose x_t and Y_t are both Granger caused by Z_t. x_t is likely to Granger cause Y_t in a model that omits Z_t if $\mathrm{E}[Y_t|Y_{t-1},x_{t-1},\ldots] \neq \mathrm{E}[Y_t|Y_{t-1},\ldots]$ even though $\mathrm{E}[Y_t|Y_{t-1},Z_{t-1},x_{t-1},\ldots] = \mathrm{E}[Y_t|Y_{t-1},Z_{t-1},\ldots]$.

Testing

Testing Granger causality in a VAR(P) is implemented using a likelihood ratio test. In the VAR(P),

$$\mathbf{Y}_t = \Phi_0 + \Phi_1\mathbf{Y}_{t-1} + \Phi_2\mathbf{Y}_{t-2} + \ldots + \Phi_P\mathbf{Y}_{t-P} + \varepsilon_t,$$

$Y_{j,t}$ does not Granger cause $Y_{i,t}$ if $\phi_{ij,1} = \phi_{ij,2} = \ldots = \phi_{ij,P} = 0$. The likelihood ratio test statistic for testing the null $H_0 : \phi_{ij,m} = 0, \ \forall m \in \{1,2,\ldots,P\}$ against the alternative $H_1 : \phi_{ij,m} \neq 0 \ \exists m \in \{1,2,\ldots,P\}$ is

$$\left(T - (Pk^2 - k)\right)\left(\ln|\hat{\Sigma}_r| - \ln|\hat{\Sigma}_u|\right) \overset{A}{\sim} \chi^2_P$$

where Σ_r is the estimated residual covariance when the null of no Granger causation is imposed ($H_0 : \phi_{ij,1} = \phi_{ij,2} = \ldots = \phi_{ij,P} = 0$) and Σ_u is the estimated covariance in the unrestricted VAR(P).[12]

5.6.1 Example: Monetary Policy VAR

The monetary policy VAR is used to illustrate testing Granger causality. Table 5.4 contains the results of Granger causality tests in the monetary policy VAR with three lags (as chosen by the HQIC). Tests of a variable causing itself have been omitted since these are not informative about the need for a

---

[11]Technically, this definition is for Granger causality in the mean. Other definition exist for Granger causality in the variance (replace conditional expectation with conditional variance) and distribution (replace conditional expectation with conditional distribution).

[12]The multiplier in the test is a degree of freedom adjusted factor. There are T data points, and there are $Pk^2 - k$ parameters in the restricted model.

| | Fed. Funds Rate | | Inflation | | Unemployment | |
Exclusion	P-val	Stat	P-val	Stat	P-val	Stat
Fed. Funds Rate	–	–	0.001	13.068	0.014	8.560
Inflation	0.001	14.756	–	–	0.375	1.963
Unemployment	0.000	19.586	0.775	0.509	–	–
All	0.000	33.139	0.000	18.630	0.005	10.472

Table 5.4: Tests of Granger causality. This table contains tests where the variable on the left-hand side is excluded from the regression for the variable along the top. Since the null is no GC, rejection indicates a relationship between past values of the variable on the left and contemporaneous values of variables on the top.

multivariate model. The table contains tests whether the variables in the left-hand column Granger Cause the variables labeled across the top. Each row contains a p-value indicating significance using standard test sizes (5 or 10%), and so each variable causes at least one other variable. Column-by-column examination demonstrated that every variable is caused by at least one other variable. The final row labeled All tests the null that a univariate model performs as well as a multivariate model by restricting all variable other than the target to have zero coefficients. This test further confirms that the VAR is required for each component.

5.7   Impulse Response Functions

In the univariate world, the $MA(\infty)$ representation of an ARMA is sufficient to understand how a shock decays. When analyzing vector data, this is no longer the case. A shock to one series has an immediate effect on that series, but it can also affect the other variables in the system which, in turn, feed back into the original variable. It is not possible to visualize the propagation of a shock using only the estimated parameters in a VAR. Impulse response functions simplify this task by providing a visual representation of shock propagation.

5.7.1   Defined

Definition 5.12 (Impulse Response Function)**.** The impulse response function of Y_i, an element of $\mathbf{Y}$, with respect to a shock in ε_j, an element of ε, for any j and i, is defined as the change in Y_{it+s}, $s \geq 0$ for a one standard deviation shock in $\varepsilon_{j,t}$.

This definition is somewhat difficult to parse and the impulse response function easier to understand using the vector moving average (VMA) representation of a VAR.[13] When $\mathbf{Y}_t$ is covariance stationary then it must have a VMA representation,

$$\mathbf{Y}_t = \mu + \varepsilon_t + \Xi_1 \varepsilon_{t-1} + \Xi_2 \varepsilon_{t-2} + \dots$$

---

[13]Recall that a stationary AR(P) can also be transformed into a $MA(\infty)$. Transforming a stationary VAR(P) into a VMA(∞) is the multivariate time-series analogue.

Using this VMA, the impulse response of Y_i with respect to a shock in ε_j at period h is

$$IRF_h = \sigma_j \Xi_h \mathbf{e}_j \qquad (5.25)$$

where $\mathbf{e}_j$ is a vector of 0s with 1 in position j, $\mathbf{e}_j = \left[\underbrace{0,\ldots,0}_{j-1}, 1, \underbrace{0,\ldots,0}_{k-j} \right]'$ and where σ_j is the stan-

dard deviation of ε_j. These impulse responses are then $\left\{ \sigma_j, \sigma_j \Xi_1^{[ii]}, \sigma_j \Xi_2^{[ii]}, \sigma_j \Xi_3^{[ii]}, \ldots \right\}$ if $i = j$ and
$\left\{ 0, \sigma_j \Xi_1^{[ij]}, \sigma_j \Xi_2^{[ij]}, \sigma_j \Xi_3^{[ij]}, \ldots \right\}$ otherwise where $\Xi_m^{[ij]}$ is the element in row i and column j of Ξ_m. The
coefficients of the VMA can be computed from the VAR using the relationship

$$\Xi_j = \Phi_1 \Xi_{j-1} + \Phi_2 \Xi_{j-2} + \ldots + \Phi_P \Xi_{j-P}$$

where $\Xi_0 = \mathbf{I}_k$ and $\Xi_m = \mathbf{0}$ for $m < 0$. For example, in a VAR(2),

$$\mathbf{Y}_t = \Phi_1 \mathbf{Y}_{t-1} + \Phi_2 \mathbf{Y}_{t-2} + \varepsilon_t,$$

$\Xi_0 = \mathbf{I}_k$, $\Xi_1 = \Phi_1$, $\Xi_2 = \Phi_1^2 + \Phi_2$, and $\Xi_3 = \Phi_1^3 + \Phi_1\Phi_2 + \Phi_2\Phi_1$.

5.7.2 Orthogonal Impulse Response Functions

The previous discussion assumed shocks are uncorrelated so that a shock to component j had no
effect on the other components of the error. This assumption is problematic since the shocks are
often correlated, and so it is not possible to change one in isolation. The model shocks have covari-
ance $\text{Cov}[\varepsilon_t] = \Sigma$, and so a set of orthogonal shocks can be produced as $\eta_t = \Sigma^{-1/2}\varepsilon_t$. Using these
uncorrelated and standardized shocks, the VMA is now

$$\mathbf{Y}_t = \mu + \varepsilon_t + \Xi_1 \Sigma^{1/2} \Sigma^{-1/2} \varepsilon_{t-1} + \Xi_2 \Sigma^{1/2} \Sigma^{-1/2} \varepsilon_{t-2} + \ldots$$
$$= \mu + \Sigma^{1/2}\eta_t + \tilde{\Xi}_1 \eta_{t-1} + \tilde{\Xi}_2 \eta_{t-2} + \ldots$$

where $\tilde{\Xi}_m = \Xi_m \Sigma^{1/2}$. The impulse response for a shock to series j in period h is $\Sigma^{1/2}\mathbf{e}_j$ in period 0,

$$OIRF_h = \tilde{\Xi}_h \mathbf{e}_j \qquad (5.26)$$

for $h \geq 1$. If Σ is diagonal, then these impulse responses are identical to the expression in eq. (5.25).

In practice, the Cholesky factor is used as the square root of the covariance matrix. The Cholesky
factor is a lower triangular matrix which imposes a de facto ordering to the shocks. For example, if

$$\Sigma = \begin{bmatrix} 1 & 1 \\ 1 & 4 \end{bmatrix},$$

then the Cholesky factor is

$$\Sigma_C^{1/2} = \begin{bmatrix} 1 & 0 \\ 1 & 2 \end{bmatrix}$$

so that $\Sigma = \Sigma_C^{1/2} \left(\Sigma_C^{1/2} \right)'$. Shocking element j has an effect of every series the appears after j $(j,\ldots,k)$
but not on the first $j-1$ $(1,\ldots,j-1)$. In some contexts, it is plausible that there is a natural order to
the shocks since some series are faster than others.

In the monetary policy VAR, it is commonly assumed that changes in the Federal Funds rate immediately spillover to unemployment and inflation, but that unemployment and inflation only feed-back into the Federal Funds rate with a lag. Similarly, it is commonly assumed that changes in unemployment affect inflation immediately, but that inflation does not have a contemporaneous impact on unemployment. When using the Cholesky factor, the impulse responses depend on the order of the variables in the VAR. Additionally, in many important applications – for example when a VAR includes multiple financial variables – then there is no plausible method to order the shocks since financial variables are likely to react simultaneously to a shock.

The leading alternative to the using the Cholesky factor is to use a Generalized Impulse Response function (Pesaran and Shin, 1998). This method is invariant to the order of the variables since it does not use a matrix square root. The GIRF is justified as the difference measuring between the conditional expectation of $\mathbf{Y}_{t+h}$ given shock j is one standard deviation and the conditional expectation of $\mathbf{Y}_{t+h}$,

$$\mathrm{E}_t\left[\mathbf{Y}_{t+h}|\varepsilon_j = \sigma_j\right] - \mathrm{E}_t\left[\mathbf{Y}_{t+h}\right].$$

When the VAR is driven by normally distributed errors, this expression is

$$GIRF_h = \sigma_j^{-1}\Xi_h\Sigma\mathbf{e}_j. \tag{5.27}$$

The GIRF is equivalently expressed as

$$\Xi_h\left[\sigma_{1j}, \sigma_{2j}, \ldots, \sigma_{kj}\right]'/\sigma_{jj} \times \sigma_j = \Xi_h\left[\beta_{1j}, \beta_{2j}, \ldots, \beta_{kj}\right]'\sigma_j$$

where β_{ij} is the population value of th regression coefficient of regressing ε_{it} on ε_{jt}.

5.7.3 Example: Impulse Response in the Monetary Policy VAR

The monetary policy VAR is used to illustrate impulse response functions. Figure 5.4 contains the impulse responses of the three variable to the three shocks. The dotted lines represent two standard deviation confidence intervals. The covariance is factored using the Cholesky, and it is assumed that the shock to the Federal Funds Rate impacts all variables immediately, the shock the unemployment affects inflation immediately but not the Federal Funds rate, and that the inflation shock has no immediate effect. The unemployment rate is sensitive to changes in the Federal Funds rate, and one standard deviation shock reduces the change ($\Delta\mathrm{UNEMP}_t$) in the unemployment rate by up to 0.15% as the impulse evolves.

5.7.4 Confidence Intervals

Impulse response functions, like the parameters of the VAR, are estimated quantities and subject to statistical variation. Confidence bands are used to determine whether an impulse response different from zero. Since the parameters of the VAR are asymptotically normally distributed (as long as it is stationary and the innovations are white noise), the impulse responses also asymptotically normal, which follows as an application of the δ-method. The analytical derivation of the covariance of the impulse response function is tedious (see Section 11.7 in Hamilton (1994) for details). Instead, two computational methods to construct confidence bands of impulse response functions are described: Monte Carlo and bootstrap.
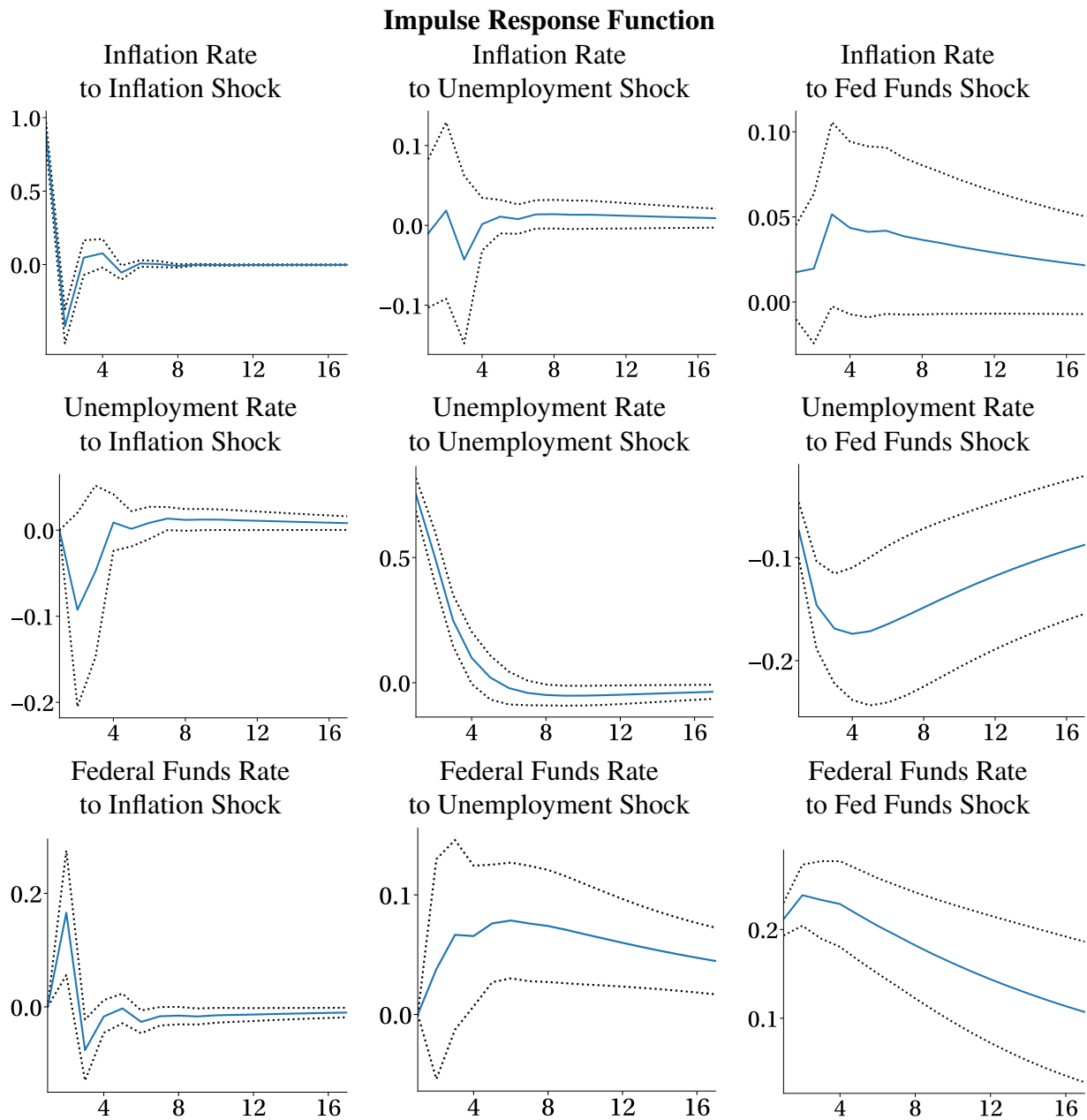
Impulse Response Function

Figure 5.4: Impulse response functions for 16 quarters. The dotted lines represent two standard deviation confidence intervals. The covariance is factored using the Cholesky so that a shock to the Federal Funds rate spills over immediately to the other two variables, an unemployment shock spills over to inflation, and an inflation shock has no immediate effect on the other series.

5.7.4.1 Monte Carlo Confidence Intervals

Monte Carlo confidence intervals come in two forms, one that directly simulates $\hat{\Phi}_i$ from its asymptotic distribution and one that simulates the VAR and draws $\hat{\Phi}_i$ as the result of estimating the unknown parameters in the simulated VAR. The direct sampling method is simple:

1. Compute $\hat{\theta}$ from the data and estimate the covariance matrix $\hat{\Lambda}$ in the asymptotic distribution $\sqrt{T}(\hat{\theta} - \theta) \overset{A}{\sim} N(0, \hat{\Lambda})$ where θ is the collection of all model parameters, $\Phi_0, \Phi_1, \ldots, \Phi_P$ and Σ.

2. Using $\hat{\theta}$ and $\hat{\Lambda}$, generate simulated values $\hat{\Phi}_{0b}, \hat{\Phi}_{1b}, \ldots, \hat{\Phi}_{Pb}$ and $\hat{\Sigma}_b$ from the asymptotic distribution as $\hat{\theta} + \hat{\Lambda}^{1/2} \varepsilon$ where $\varepsilon \overset{\text{i.i.d.}}{\sim} N(\mathbf{0}, \mathbf{I}_{k^2(P+1)})$. These are i.i.d. draws from a $N(\hat{\theta}, \hat{\Lambda})$ distribution.

3. Using $\hat{\Phi}_{0b}, \hat{\Phi}_{1b}, \ldots, \hat{\Phi}_{Pb}$ and $\hat{\Sigma}_b$, compute the impulse responses $\{\hat{\Xi}_{jb}\}$ where $j = 1, 2, \ldots, h$. Save these values.

4. Return to step 2 and compute a total of B impulse responses. Typically B is between 100 and 1000.

5. For each impulse response and each horizon, sort the responses. The 5th and 95th percentile of this distribution are the confidence intervals.

The second Monte Carlo method simulates data assuming the errors are i.i.d. normally distributed, and then uses these values to produce a draw from the joint distribution of the model parameters. This method avoids the estimation of the parameter covariance matrix $\hat{\Lambda}$ in the alternative Monte Carlo method.

1. Compute $\hat{\Phi}$ from the initial data and estimate the residual covariance $\hat{\Sigma}$.

2. Using $\hat{\Phi}$ and $\hat{\Sigma}$, simulate a time-series $\{\tilde{\mathbf{Y}}_t\}$ with as many observations as the original data. These can be computed directly using forward recursion

$$\tilde{\mathbf{Y}}_t = \hat{\Phi}_0 + \hat{\Phi}_1 \mathbf{Y}_{t-1} + \ldots + \hat{\Phi}_P \mathbf{Y}_{t-P} + \hat{\Sigma}^{1/2} \varepsilon_t$$

where $\varepsilon \overset{\text{i.i.d.}}{\sim} N(\mathbf{0}, \mathbf{I}_k)$ are multivariate standard normally distributed. The P initial values are set to a consecutive block of the historical data chosen at random, $\mathbf{Y}_\tau, \mathbf{Y}_{\tau+1}, \ldots, \mathbf{Y}_{\tau+P-1}$ for $\tau \in \{1, \ldots, T-P\}$.

3. Using $\{\tilde{\mathbf{Y}}_t\}$, estimate the model parameters $\hat{\Phi}_{0b}, \hat{\Phi}_{1b}, \ldots, \hat{\Phi}_{Pb}$ and $\hat{\Sigma}_b$.

4. Using $\hat{\Phi}_{0b}, \hat{\Phi}_{1b}, \ldots, \hat{\Phi}_{Pb}$ and $\hat{\Sigma}_b$, compute the impulse responses $\{\hat{\Xi}_{jb}\}$ where $j = 1, 2, \ldots, h$. Save these values.

5. Return to step 2 and compute a total of B impulse responses. Typically B is between 100 and 1000.

6. For each impulse response for each horizon, sort the impulse responses. The 5th and 95th percentile of this distribution are the confidence intervals.

Of these two methods, the former should be preferred since the assumption of i.i.d. normally distributed errors in the latter may be unrealistic, especially when modeling financial data.

5.7.4.2   Bootstrap Confidence Intervals

The bootstrap is a simulation-based method that resamples from the observed data produce a simulated data set. The idea behind this method is simple: if the residuals are realizations of the actual error process, one can use them directly to simulate this distribution rather than making an arbitrary assumption about the error distribution (e.g., i.i.d. normal). The procedure is essentially identical to the second Monte Carlo procedure outlined above:

1. Compute $\hat{\Phi}$ from the initial data and estimate the residuals $\hat{\varepsilon}_t$.

2. Using $\hat{\varepsilon}_t$, compute a new series of residuals $\tilde{\varepsilon}_t$ by sampling, with replacement, from the original residuals. The new series of residuals can be described

$$\{\hat{\varepsilon}_{u_1}, \hat{\varepsilon}_{u_2}, \ldots, \hat{\varepsilon}_{u_T}\}$$

   where u_i are i.i.d. discrete uniform random variables taking the values $1, 2, \ldots, T$. In essence, the new set of residuals is just the old set of residuals reordered with some duplication and omission.[14]

3. Using $\hat{\Phi}$ and $\{\hat{\varepsilon}_{u_1}, \hat{\varepsilon}_{u_2}, \ldots, \hat{\varepsilon}_{u_T}\}$, simulate a time-series $\{\tilde{\mathbf{Y}}_t\}$ with as many observations as the original data. These can be computed directly using the VAR

$$\tilde{\mathbf{Y}}_t = \hat{\Phi}_0 + \hat{\Phi}_1 \mathbf{Y}_{t-1} + \ldots + \hat{\Phi}_P \mathbf{Y}_{t-P} + \hat{\varepsilon}_{u_t}$$

4. Using $\{\tilde{\mathbf{Y}}_t\}$, compute estimates of $\hat{\Phi}_{0b}, \hat{\Phi}_{1b}, \ldots, \hat{\Phi}_{Pb}$ and $\hat{\Sigma}_b$ from a VAR.

5. Using $\hat{\Phi}_{0b}, \hat{\Phi}_{1b}, \ldots, \hat{\Phi}_{Pb}$ and $\hat{\Sigma}_b$, compute the impulse responses $\{\breve{\Xi}_{jb}\}$ where $j = 1, 2, \ldots, h$. Save these values.

6. Return to step 2 and compute a total of B impulse responses. Typically B is between 100 and 1000.

7. For each impulse response for each horizon, sort the impulse responses. The 5th and 95th percentile of this distribution are the confidence intervals.

5.8   Cointegration

Many economic time-series are nonstationarity and so standard VAR analysis which assumes all series are covariance stationary is unsuitable. Cointegration extends stationary VAR models to non-stationary time series. Cointegration analysis also provides a method to characterize the long-run equilibrium of a system of non-stationary variables. Before more formally examining cointegration, consider the consequences if two economic variables that have been widely documented to contain unit roots, consumption and income, have no long-run relationship. Without a stable equilibrium relationship, the values of these two variables would diverge over time. Individuals would either have extremely high saving rates – when income is far above consumption, or become incredibly indebted.

---

[14]This is one version of the bootstrap and is appropriate for homoskedastic data. If the data are heteroskedastic, some form of block bootstrap is needed.

These two scenarios are implausible, and so there must be some long-run (or equilibrium) relationship between consumption and income. Similarly, consider the relationship between the spot and future price of oil. Standard finance theory dictates that future's price, f_t, is a conditionally unbiased estimate of the spot price in period $t + 1$, s_{t+1} ($E_t[s_{t+1}] = f_t$, assuming various costs such as the risk-free rate and storage are 0). Additionally, today's spot price is also an unbiased estimate of tomorrow's spot price ($E_t[s_{t+1}] = s_t$). However, both the spot and future price contain unit roots. Combining these two identities reveals a cointegrating relationship: $s_t - f_t$ should be stationary even if the spot and future prices contain unit roots.[15]

In stationary time-series, whether scalar or when the multiple processes are linked through a VAR, the process is self-equilibrating; given enough time, a process reverts to its unconditional mean. In a VAR, both the individual series and linear combinations of the series are stationary. The behavior of cointegrated processes is meaningfully different. Each component of a cointegrated process contains a unit root, and so has shocks with a permanent impact. However, when combined with another series, a cointegrated pair revert towards one another. A cointegrated pair is mean reverting to a stochastic trend (a unit root process), rather than to fixed value.

Cointegration and error correction provide a set of tools to analyze long-run relationships and short-term deviations from the equilibria. Cointegrated time-series exhibit temporary deviations from a long-run trend but are ultimately mean reverting to this trend. The Vector Error Correction Model (VECM) explicitly includes the deviation from the long-run relationship when modeling the short-term dynamics of the time series to push the components towards their long-run relationship.

5.8.1   Definition

Recall that a first-order integrated process is not stationary in levels but is stationary in differences. When this is the case, Y_t is $I(1)$ and $\Delta Y_t = Y_t - Y_{t-1}$ is $I(0)$. Cointegration builds on this structure by defining relationships across series which transform multiple $I(1)$ series into $I(0)$ series without using time-series differences.

Definition 5.13 (Bivariate Cointegration). Let $\{x_t\}$ and $\{Y_t\}$ be two $I(1)$ series. These series are cointegrated if there exists a vector β with both elements non-zero such that

$$\beta'[x_t \ Y_t]' = \beta_1 x_t - \beta_2 Y_t \sim I(0) \tag{5.28}$$

This definition states that there exists a nontrivial linear combination of x_t and Y_t that is stationary. This feature – a stable relationship between the two series, is a powerful tool in the analysis of nonstationary data. When treated individually, the data are extremely persistent; however, there is a well-behaved linear combination with transitory shocks that is stationary. Moreover, in many cases, this relationship takes a meaningful form such as $Y_t - x_t$.

Cointegrating relationships are only defined up to a non-zero constant. For example if $x_t - \beta Y_t$ is a cointegrating relationship, then $2x_t - 2\beta Y_t = 2(x_t - \beta Y_t)$ is also a cointegrating relationship. The standard practice is to normalize the vector on one of the variables so that its coefficient is unity. For example, if $\beta_1 x_t - \beta_2 Y_t$ is a cointegrating relationship, the two normalized versions are $x_t - \beta_2/\beta_1 Y_t = x_t - \tilde{\beta} Y_t$ and $Y_t - \beta_1/\beta_2 x_t = Y_t - \ddot{\beta} x_t$.

The complete definition in the general case is similar, albeit slightly more intimidating.

---

[15]This assumes the horizon is short.

Nonstationary and Stationary VAR(1)s

Cointegration (Φ_{11}) Independent Unit Roots(Φ_{12})

Persistent, Stationary (Φ_{21}) Anti-persistent, Stationary(Φ_{22})
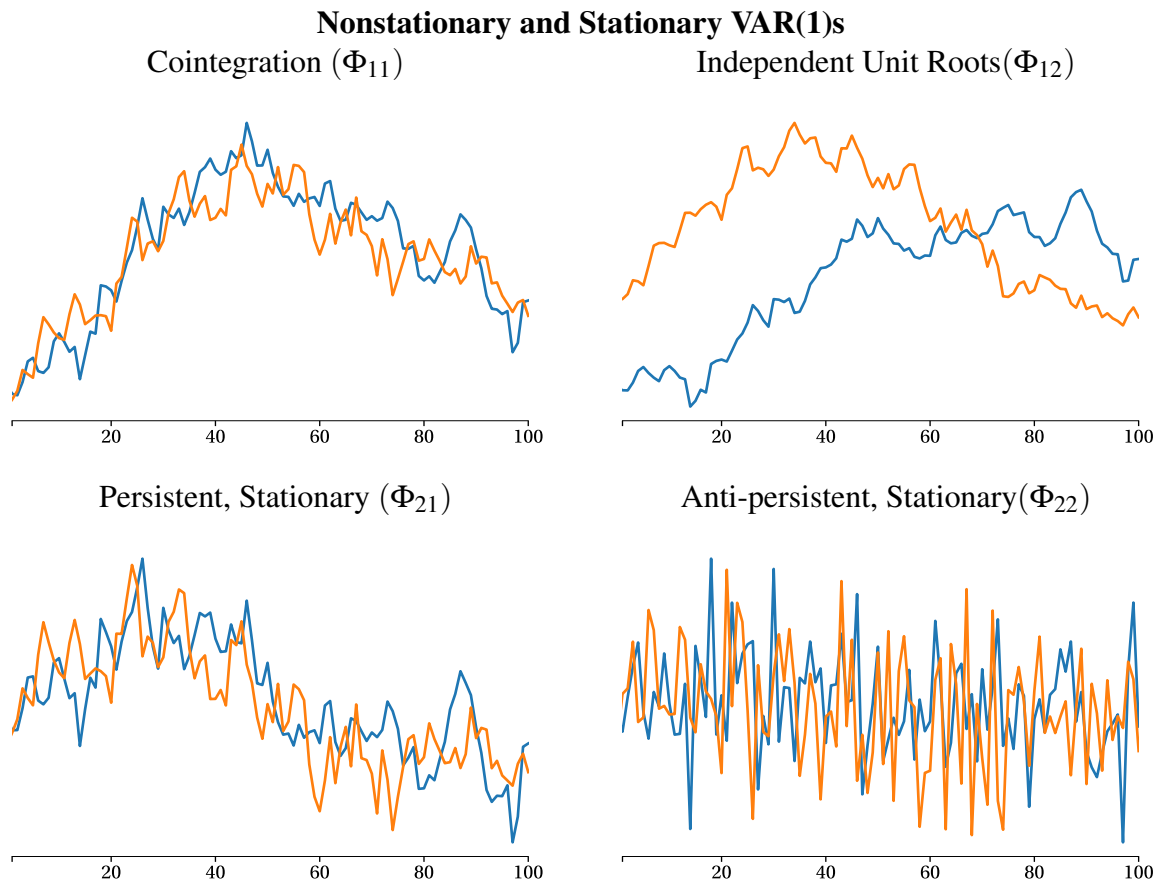
Figure 5.5: A plot of four time-series that all begin at the same initial value and use the same shocks. All data are generated by $\mathbf{Y}_t = \Phi_{ij}\mathbf{Y}_{t-1} + \varepsilon_t$ where Φ_{ij} varies across the panels.

Definition 5.14 (Cointegration). A set of k variables $\mathbf{Y}_t$ are cointegrated if at least two series are $I(1)$ and there exists a non-zero, reduced rank k by k matrix π such that

$$\pi\mathbf{Y}_t \sim I(0). \tag{5.29}$$

The non-zero requirement is obvious: if $\pi = \mathbf{0}$ then $\pi\mathbf{Y}_t = \mathbf{0}$ and this time series is trivially $I(0)$. The second requirement that π is reduced rank is not. This technical requirement is necessary since whenever π is full rank and $\pi\mathbf{Y}_t \sim I(0)$, the series must be the case that $\mathbf{Y}_t$ is also $I(0)$. However, for variables to be *cointegrated,* they must be *integrated*. If the matrix is full rank, the common unit roots cannot cancel, and $\pi\mathbf{Y}_t$ must have the same order of integration as $\mathbf{Y}$. Finally, the requirement that at least two of the series are $I(1)$ rules out the degenerate case where all components of $\mathbf{Y}_t$ are $I(0)$, and allows $\mathbf{Y}_t$ to contain both $I(0)$ and $I(1)$ random variables. If $\mathbf{Y}_t$ contains both $I(0)$ and $I(1)$ random variables, then the long-run relationship only depends on the $I(1)$ random variable.

For example, suppose the components of $\mathbf{Y}_t = [Y_{1t}, Y_{2t}]'$ are cointegrated so that $Y_{1t} - \beta Y_{2t}$ is stationary. One choice for π is

$$\pi = \begin{bmatrix} 1 & -\beta \\ 1 & -\beta \end{bmatrix}$$

To begin developing an understanding of cointegration, examine the plots in Figure 5.5. These four plots show two nonstationary processes and two stationary processes *all initialized at the same value and using the same shocks*. These plots contain simulated data from VAR(1) processes with different parameters, Φ_{ij}.

$$\mathbf{Y}_t = \Phi_{ij}\mathbf{Y}_{t-1} + \varepsilon_t,$$

$$\Phi_{11} = \begin{bmatrix} .8 & .2 \\ .2 & .8 \end{bmatrix}, \qquad \Phi_{12} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix},$$
$$\lambda_i = 1, 0.6 \qquad\qquad \lambda_i = 1, 1$$

$$\Phi_{21} = \begin{bmatrix} .7 & .2 \\ .2 & .7 \end{bmatrix}, \qquad \Phi_{22} = \begin{bmatrix} -.3 & .3 \\ .1 & -.2 \end{bmatrix},$$
$$\lambda_i = 0.9, 0.5 \qquad\qquad \lambda_i = -0.43, -0.06$$

where λ_i are the eigenvalues of the parameter matrices. The nonstationary processes both have unit eigenvalues. The eigenvalues in the stationary processes are all less than 1 (in absolute value). The cointegrated process has a single unit eigenvalue while the independent unit root process has two. In a VAR(1), the number of unit eigenvalues plays a crucial role in cointegration and higher dimension cointegrated systems may contain between 1 and $k-1$ unit eigenvalues. The number of unit eigenvalues shows the count of the unit root "drivers" in the system of equations.[16] The picture presents evidence of the most significant challenge in cointegration analysis: it can be challenging to tell when two series are cointegrated, a feature in common with unit root testing of a single time series.

5.8.2   Vector Error Correction Models (VECM)

The Granger representation theorem provides a key insight into cointegrating relationships. Granger demonstrated that if a system is cointegrated then there exists a vector error correction model with a reduced rank coefficient matrix and if there is a VECM with a reduced rank coefficient matrix then the system must be cointegrated. A VECM describes the short-term deviations from the long-run trend (a stochastic trend/unit root). The simplest VECM is

$$\begin{bmatrix} \Delta x_t \\ \Delta Y_t \end{bmatrix} = \begin{bmatrix} \pi_{11} & \pi_{12} \\ \pi_{21} & \pi_{22} \end{bmatrix} \begin{bmatrix} x_{t-1} \\ Y_{t-1} \end{bmatrix} + \begin{bmatrix} \varepsilon_{1,t} \\ \varepsilon_{2,t} \end{bmatrix} \tag{5.30}$$

which states that changes in x_t and Y_t are related to the levels of x_t and Y_t through the cointegrating matrix (π). However, since x_t and Y_t are cointegrated, there exists β such that $x_t - \beta Y_t = \begin{bmatrix} 1 & -\beta \end{bmatrix} \begin{bmatrix} x_t & Y_t \end{bmatrix} \sim I(0)$. Substituting this value into this equation, equation 5.30 is equivalently expressed as

$$\begin{bmatrix} \Delta x_t \\ \Delta Y_t \end{bmatrix} = \begin{bmatrix} \alpha_1 \\ \alpha_2 \end{bmatrix} \begin{bmatrix} 1 & -\beta \end{bmatrix} \begin{bmatrix} x_{t-1} \\ Y_{t-1} \end{bmatrix} + \begin{bmatrix} \varepsilon_{1,t} \\ \varepsilon_{2,t} \end{bmatrix}. \tag{5.31}$$

The short-run dynamics evolve according to

$$\Delta x_t = \alpha_1(x_{t-1} - \beta Y_{t-1}) + \varepsilon_{1,t} \tag{5.32}$$

---

[16]In higher order VAR models, the eigenvalues must be computed from the companion form.

and

$$\Delta Y_t = \alpha_2(x_{t-1} - \beta Y_{t-1}) + \varepsilon_{2,t}. \tag{5.33}$$

The important elements of this VECM can be clearly labeled: $x_{t-1} - \beta Y_{t-1}$ is the deviation from the long-run trend (also known as the equilibrium correction term) and α_1 and α_2 are the speed of adjustment parameters. VECMs impose one restriction of the αs: they cannot both be 0 (if they were, π would also be $\mathbf{0}$). In its general form, an VECM can be augmented to allow past short-run deviations to also influence present short-run deviations and to include deterministic trends. In vector form, an VECM(P) evolves according to

$$\Delta \mathbf{Y}_t = \delta_0 + \pi \mathbf{Y}_{t-1} + \pi_1 \Delta \mathbf{Y}_{t-1} + \pi_2 \Delta \mathbf{Y}_{t-2} + \ldots + + \pi_P \Delta \mathbf{Y}_{t-P} + \varepsilon_t$$

where $\pi \mathbf{Y}_{t-1} = \alpha \beta' \mathbf{Y}_t$ captures the cointegrating relationship, δ_0 represents a linear time trend in the original data (levels) and $\pi_j \Delta \mathbf{Y}_{t-j}$, $j = 1, 2, \ldots, P$ capture short-run dynamics around the stochastic trend.

5.8.2.1 The Mechanics of the VECM

Any cointegrated VAR can be transformed into an VECM. Consider a simple cointegrated bivariate VAR(1)

$$\begin{bmatrix} x_t \\ Y_t \end{bmatrix} = \begin{bmatrix} .8 & .2 \\ .2 & .8 \end{bmatrix} \begin{bmatrix} x_{t-1} \\ Y_{t-1} \end{bmatrix} + \begin{bmatrix} \varepsilon_{1,t} \\ \varepsilon_{2,t} \end{bmatrix}$$

To transform this VAR to an VECM, begin by subtracting $[x_{t-1} \ Y_{t-1}]'$ from both sides

$$\begin{bmatrix} x_t \\ Y_t \end{bmatrix} - \begin{bmatrix} x_{t-1} \\ Y_{t-1} \end{bmatrix} = \begin{bmatrix} .8 & .2 \\ .2 & .8 \end{bmatrix} \begin{bmatrix} x_{t-1} \\ Y_{t-1} \end{bmatrix} - \begin{bmatrix} x_{t-1} \\ Y_{t-1} \end{bmatrix} + \begin{bmatrix} \varepsilon_{1,t} \\ \varepsilon_{2,t} \end{bmatrix} \tag{5.34}$$

$$\begin{bmatrix} \Delta x_t \\ \Delta Y_t \end{bmatrix} = \left(\begin{bmatrix} .8 & .2 \\ .2 & .8 \end{bmatrix} - \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \right) \begin{bmatrix} x_{t-1} \\ Y_{t-1} \end{bmatrix} + \begin{bmatrix} \varepsilon_{1,t} \\ \varepsilon_{2,t} \end{bmatrix}$$

$$\begin{bmatrix} \Delta x_t \\ \Delta Y_t \end{bmatrix} = \begin{bmatrix} -.2 & .2 \\ .2 & -.2 \end{bmatrix} \begin{bmatrix} x_{t-1} \\ Y_{t-1} \end{bmatrix} + \begin{bmatrix} \varepsilon_{1,t} \\ \varepsilon_{2,t} \end{bmatrix}$$

$$\begin{bmatrix} \Delta x_t \\ \Delta Y_t \end{bmatrix} = \begin{bmatrix} -.2 \\ .2 \end{bmatrix} \begin{bmatrix} 1 & -1 \end{bmatrix} \begin{bmatrix} x_{t-1} \\ Y_{t-1} \end{bmatrix} + \begin{bmatrix} \varepsilon_{1,t} \\ \varepsilon_{2,t} \end{bmatrix}$$

In this example, the speed of adjustment parameters are $-.2$ for Δx_t and $.2$ for ΔY_t and the normalized (on x_t) cointegrating relationship is $[1 \ -1]$.

In the general multivariate case, a cointegrated VAR(P) can be turned into an VECM by recursive substitution. Consider a cointegrated VAR(3),

$$\mathbf{Y}_t = \Phi_1 \mathbf{Y}_{t-1} + \Phi_2 \mathbf{Y}_{t-2} + \Phi_3 \mathbf{Y}_{t-3} + \varepsilon_t$$

This system is cointegrated if at least one but fewer than k eigenvalues of $\pi = \Phi_1 + \Phi_2 + \Phi_3 - \mathbf{I}_k$ are not zero. To begin the transformation, add and subtract $\Phi_3 \mathbf{Y}_{t-2}$ to the right side

$$\begin{aligned}
\mathbf{Y}_t &= \Phi_1\mathbf{Y}_{t-1} + \Phi_2\mathbf{Y}_{t-2} + \Phi_3\mathbf{Y}_{t-2} - \Phi_3\mathbf{Y}_{t-2} + \Phi_3\mathbf{Y}_{t-3} + \varepsilon_t \\
&= \Phi_1\mathbf{Y}_{t-1} + \Phi_2\mathbf{Y}_{t-2} + \Phi_3\mathbf{Y}_{t-2} - \Phi_3\Delta\mathbf{Y}_{t-2} + \varepsilon_t \\
&= \Phi_1\mathbf{Y}_{t-1} + (\Phi_2 + \Phi_3)\mathbf{Y}_{t-2} - \Phi_3\Delta\mathbf{Y}_{t-2} + \varepsilon_t.
\end{aligned}$$

Next, add and subtract $(\Phi_2 + \Phi_3)\mathbf{Y}_{t-1}$ to the right-hand side,

$$\begin{aligned}
\mathbf{Y}_t &= \Phi_1\mathbf{Y}_{t-1} + (\Phi_2 + \Phi_3)\mathbf{Y}_{t-1} - (\Phi_2 + \Phi_3)\mathbf{Y}_{t-1} + (\Phi_2 + \Phi_3)\mathbf{Y}_{t-2} - \Phi_3\Delta\mathbf{Y}_{t-2} + \varepsilon_t \\
&= \Phi_1\mathbf{Y}_{t-1} + (\Phi_2 + \Phi_3)\mathbf{Y}_{t-1} - (\Phi_2 + \Phi_3)\Delta\mathbf{Y}_{t-1} - \Phi_3\Delta\mathbf{Y}_{t-2} + \varepsilon_t \\
&= (\Phi_1 + \Phi_2 + \Phi_3)\mathbf{Y}_{t-1} - (\Phi_2 + \Phi_3)\Delta\mathbf{Y}_{t-1} - \Phi_3\Delta\mathbf{Y}_{t-2} + \varepsilon_t.
\end{aligned}$$

Finally, subtract $\mathbf{Y}_{t-1}$ from both sides,

$$\begin{aligned}
\mathbf{Y}_t - \mathbf{Y}_{t-1} &= (\Phi_1 + \Phi_2 + \Phi_3)\mathbf{Y}_{t-1} - \mathbf{Y}_{t-1} - (\Phi_2 + \Phi_3)\Delta\mathbf{Y}_{t-1} - \Phi_3\Delta\mathbf{Y}_{t-2} + \varepsilon_t \\
\Delta\mathbf{Y}_t &= (\Phi_1 + \Phi_2 + \Phi_3 - \mathbf{I}_k)\mathbf{Y}_{t-1} - (\Phi_2 + \Phi_3)\Delta\mathbf{Y}_{t-1} - \Phi_3\Delta\mathbf{Y}_{t-2} + \varepsilon_t.
\end{aligned}$$

The final step is to relabel the equation in terms of π notation,

$$\begin{aligned}
\mathbf{Y}_t - \mathbf{Y}_{t-1} &= (\Phi_1 + \Phi_2 + \Phi_3 - \mathbf{I}_k)\mathbf{Y}_{t-1} - (\Phi_2 + \Phi_3)\Delta\mathbf{Y}_{t-1} - \Phi_3\Delta\mathbf{Y}_{t-2} + \varepsilon_t \qquad (5.35)\\
\Delta\mathbf{Y}_t &= \pi\mathbf{Y}_{t-1} + \pi_1\Delta\mathbf{Y}_{t-1} + \pi_2\Delta\mathbf{Y}_{t-2} + \varepsilon_t.
\end{aligned}$$

which is equivalent to

$$\Delta\mathbf{Y}_t = \alpha\beta'\mathbf{Y}_{t-1} + \pi_1\Delta\mathbf{Y}_{t-1} + \pi_2\Delta\mathbf{Y}_{t-2} + \varepsilon_t. \qquad (5.36)$$

where α contains the speed of adjustment parameters, and β contains the cointegrating vectors. This recursion can be used to transform any VAR(P), whether cointegrated or not,

$$\mathbf{Y}_{t-1} = \Phi_1\mathbf{Y}_{t-1} + \Phi_2\mathbf{Y}_{t-2} + \ldots + \Phi_P\mathbf{Y}_{t-P} + \varepsilon_t$$

into its VECM from

$$\Delta\mathbf{Y}_t = \pi\mathbf{Y}_{t-1} + \pi_1\Delta\mathbf{Y}_{t-1} + \pi_2\Delta\mathbf{Y}_{t-2} + \ldots + \pi_{P-1}\Delta\mathbf{Y}_{t-P+1} + \varepsilon_t$$

using the identities $\pi = -\mathbf{I}_k + \sum_{i=1}^{P}\Phi_i$ and $\pi_p = -\sum_{i=p+1}^{P}\Phi_i$.[17]

---

[17]Stationary VAR(P) models can be written as VECM with one important difference. When $\{\mathbf{Y}_t\}$ is covariance stationary, then π must have rank k. In cointegrated VAR models, the coefficient π in the VECM always has rank between 1 and $k-1$. If π has rank 0, then the VAR(P) contains k distinct unit roots and it is note possible to construct a linear combination that is $I(0)$.

5.8.2.2 Cointegrating Vectors

The key to understanding cointegration in systems with three or more variables is to note that the matrix which governs the cointegrating relationship, π, can always be decomposed into two matrices,

$$\pi = \alpha\beta'$$

where α and β are both k by r matrices where r is the number of cointegrating relationships. For example, suppose the parameter matrix in an VECM is

$$\pi = \begin{bmatrix} 0.3 & 0.2 & -0.36 \\ 0.2 & 0.5 & -0.35 \\ -0.3 & -0.3 & 0.39 \end{bmatrix}$$

The eigenvalues of this matrix are .9758, .2142 and 0. The 0 eigenvalue of π indicates there are two cointegrating relationships since the number of cointegrating relationships is rank(π). Since there are two cointegrating relationships, β can be normalized to be

$$\beta = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ \beta_1 & \beta_2 \end{bmatrix}$$

and α has 6 unknown parameters. $\alpha\beta'$ combine to produce

$$\pi = \begin{bmatrix} \alpha_{11} & \alpha_{12} & \alpha_{11}\beta_1 + \alpha_{12}\beta_2 \\ \alpha_{21} & \alpha_{22} & \alpha_{21}\beta_1 + \alpha_{22}\beta_2 \\ \alpha_{31} & \alpha_{32} & \alpha_{31}\beta_1 + \alpha_{32}\beta_2 \end{bmatrix},$$

and α can be determined using the left block of π. Once α is known, any two of the three remaining elements can be used to solve of β_1 and β_2. Appendix A contains a detailed illustration of the steps used to find the speed of adjustment coefficients and the cointegrating vectors in trivariate cointegrated VARs.

5.8.3 Rank and the number of unit roots

The rank of π is the same as the number of distinct cointegrating vectors. Decomposing $\pi = \alpha\beta'$ shows that if π has rank r, then α and β must both have r linearly independent columns. α contains the speed of adjustment parameters, and β contains the cointegrating vectors. There are r cointegrating vectors, and so the system contains $m = k - r$ distinct unit roots. This relationship holds since when there are k variables and m distinct unit roots, it is always possible to find r distinct linear combinations eliminate the unit roots and so are stationary.

Consider a trivariate cointegrated system driven by either one or two unit roots. Denote the underlying unit root processes as $w_{1,t}$ and $w_{2,t}$. When there is a single unit root driving all three variables, the system can be expressed

$$Y_{1,t} = \kappa_1 w_{1,t} + \varepsilon_{1,t}$$
$$Y_{2,t} = \kappa_2 w_{1,t} + \varepsilon_{2,t}$$
$$Y_{3,t} = \kappa_3 w_{1,t} + \varepsilon_{3,t}$$

where $\varepsilon_{j,t}$ is a covariance stationary error (or $I(0)$, but not necessarily white noise).

In this system there are two linearly independent cointegrating vectors. First consider normalizing the coefficient on $Y_{1,t}$ to be 1 and so in the equilibrium relationship $Y_{1,t} - \beta_1 Y_{2,t} - \beta_1 Y_{3,t}$ must satisfy

$$\kappa_1 = \beta_1 \kappa_2 + \beta_2 \kappa_3.$$

This equality ensures that the unit roots are not present in the difference. This equation does not have a unique solution since there are two unknown parameters. One solution is to further restrict $\beta_1 = 0$ so that the unique solution is $\beta_2 = \kappa_1/\kappa_3$ and an equilibrium relationship is $Y_{1,t} - (\kappa_1/\kappa_3)Y_{3,t}$. This alternative normalization produces a cointegrating vector since

$$Y_{1,t} - \frac{\kappa_1}{\kappa_3}Y_{3,t} = \kappa_1 w_{1,t} + \varepsilon_{1,t} - \frac{\kappa_1}{\kappa_3}\kappa_3 w_{1,t} - \frac{\kappa_1}{\kappa_3}\varepsilon_{3,t} = \varepsilon_{1,t} - \frac{\kappa_1}{\kappa_3}\varepsilon_{3,t}$$

Alternatively one could normalize the coefficient on $Y_{2,t}$ and so the equilibrium relationship $Y_{2,t} - \beta_1 Y_{1,t} - \beta_2 Y_{3,t}$ would require

$$\kappa_2 = \beta_1 \kappa_1 + \beta_2 \kappa_3.$$

This equation is also not identified since there are two unknowns and one equation. To solve assume $\beta_1 = 0$ and so the solution is $\beta_2 = \kappa_2/\kappa_3$, which is a cointegrating relationship since

$$Y_{2,t} - \frac{\kappa_2}{\kappa_3}Y_{3,t} = \kappa_2 w_{1,t} + \varepsilon_{2,t} - \frac{\kappa_2}{\kappa_3}\kappa_3 w_{1,t} - \frac{\kappa_2}{\kappa_3}\varepsilon_{3,t} = \varepsilon_{2,t} - \frac{\kappa_2}{\kappa_3}\varepsilon_{3,t}$$

These solutions are the only two needed since any other definition of the equilibrium must be a linear combination of these. The redundant equilibrium is constructed by normalizing on $Y_{1,t}$ to define an equilibrium of the form $Y_{1,t} - \beta_1 Y_{2,t} - \beta_2 Y_{3,t}$. Imposing $\beta_3 = 0$ to identify the solution, $\beta_1 = \kappa_1/\kappa_2$ which produces the equilibrium condition

$$Y_{1,t} - \frac{\kappa_1}{\kappa_2}Y_{2,t}.$$

This equilibrium is already implied by the first two,

$$Y_{1,t} - \frac{\kappa_1}{\kappa_3}Y_{3,t} \text{ and } Y_{2,t} - \frac{\kappa_2}{\kappa_3}Y_{3,t}$$

and can be seen to be redundant since

$$Y_{1,t} - \frac{\kappa_1}{\kappa_2}Y_{2,t} = \left(Y_{1,t} - \frac{\kappa_1}{\kappa_3}Y_{3,t} \right) - \frac{\kappa_1}{\kappa_2}\left(Y_{2,t} - \frac{\kappa_2}{\kappa_3}Y_{3,t} \right)$$

In this system of three variables and one common unit root the set of cointegrating vectors can be expressed as

$$\beta = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ \frac{\kappa_1}{\kappa_3} & \frac{\kappa_2}{\kappa_3} \end{bmatrix}.$$

When a system has only one unit root and three series, there are two non-redundant linear combinations of the underlying variables which are stationary. In a complete system with k variables and a single unit root, there are $k-1$ non-redundant linear combinations that are stationary.

Next consider a trivariate system driven by two unit roots,

$$Y_{1,t} = \kappa_{11}w_{1,t} + \kappa_{12}w_{2,t} + \varepsilon_{1,t}$$
$$Y_{2,t} = \kappa_{21}w_{1,t} + \kappa_{22}w_{2,t} + \varepsilon_{2,t}$$
$$Y_{3,t} = \kappa_{31}w_{1,t} + \kappa_{32}w_{2,t} + \varepsilon_{3,t}$$

where the errors $\varepsilon_{j,t}$ are again covariance stationary but not necessarily white noise. If the coefficient on $Y_{1,t}$ is normalized to 1, then it the weights in the equilibrium condition, $Y_{1,t} - \beta_1 Y_{2,t} - \beta_2 Y_{3,t}$, satisfy

$$\kappa_{11} = \beta_1 \kappa_{21} + \beta_2 \kappa_{31}$$
$$\kappa_{12} = \beta_1 \kappa_{22} + \beta_2 \kappa_{32}$$

to order to eliminate both unit roots. This system of two equations in two unknowns has the solution

$$\begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix} = \begin{bmatrix} \kappa_{21} & \kappa_{31} \\ \kappa_{22} & \kappa_{32} \end{bmatrix}^{-1} \begin{bmatrix} \kappa_{11} \\ \kappa_{12} \end{bmatrix}.$$

This solution is unique (up to the initial normalization), and there are no other cointegrating vectors so that

$$\beta = \begin{bmatrix} 1 \\ \frac{\kappa_{11}\kappa_{32} - \kappa_{12}\kappa_{22}}{\kappa_{21}\kappa_{32} - \kappa_{22}\kappa_{31}} \\ \frac{\kappa_{12}\kappa_{21} - \kappa_{11}\kappa_{31}}{\kappa_{21}\kappa_{32} - \kappa_{22}\kappa_{31}} \end{bmatrix}$$

This line of reasoning extends to k-variate systems driven by m unit roots. One set of r cointegrating vectors is constructed by normalizing the first r elements of $\mathbf{Y}$ one at a time. In the general case

$$\mathbf{Y}_t = \mathbf{K}\mathbf{w}_t + \varepsilon_t$$

where $\mathbf{K}$ is a k by m matrix, $\mathbf{w}_t$ an m by 1 set of unit root processes, and ε_t is a k by 1 vector of covariance stationary errors. Normalizing on the first r variables, the cointegrating vectors in this system are

$$\beta = \begin{bmatrix} \mathbf{I}_r \\ \tilde{\beta} \end{bmatrix} \tag{5.37}$$

where $\mathbf{I}_r$ is an r-dimensional identity matrix. $\tilde{\beta}$ is a m by r matrix of loadings,

$$\tilde{\beta} = \mathbf{K}_2^{-1}\mathbf{K}_1', \tag{5.38}$$

where $\mathbf{K}_1$ is the first r rows of $\mathbf{K}$ (r by m) and $\mathbf{K}_2$ is the bottom m rows of $\mathbf{K}$ (m by m). In the trivariate example driven by one unit root,

$$\mathbf{K}_1 = \begin{bmatrix} \kappa_1 \\ \kappa_2 \end{bmatrix} \text{ and } \mathbf{K}_2 = \kappa_3$$

and in the trivariate system driven by two unit roots,

$$\mathbf{K}_1 = \begin{bmatrix} \kappa_{11} & \kappa_{12} \end{bmatrix} \text{ and } \mathbf{K}_2 = \begin{bmatrix} \kappa_{21} & \kappa_{22} \\ \kappa_{31} & \kappa_{32} \end{bmatrix}.$$

Applying eqs. (5.37) and (5.38) produces the previously derived set of cointegrating vectors. Note that when $r = 0$ then the system contains k unit roots and so is not cointegrated (in general) since the system would have three equations and only two unknowns. Similarly when $r = k$ there are no unit roots and any linear combination is stationary.

5.8.3.1 Relationship to Common Features and common trends

Cointegration is a particular case of a broader concept known as common features. In the case of cointegration, both series have a common stochastic trend (or common unit root). Other examples of common features include common heteroskedasticity, defined as x_t and Y_t are heteroskedastic but there exists a combination, $x_t - \beta Y_t$, which is not, common nonlinearities which are defined analogously (replacing heteroskedasticity with nonlinearity), and cobreaks, where two series both contain structural breaks but $x_t - \beta Y_t$ does now. Incorporating common features often produces simpler models than leaving them unmodeled.

5.8.4 Testing

Testing for cointegration, like testing for a unit root in a single series, is complicated. Two methods are presented, the original Engle-Granger 2-step procedure and the more sophisticated Johansen methodology. The Engle-Granger method is generally only applicable if there are two variables, if the system contains exactly one cointegrating relationship, or if the cointegration vector is known (e.g., an accounting identity where the left-hand side has to add up to the right-hand side). The Johansen methodology is substantially more general and can be used to examine complex systems with many variables and multiple cointegrating relationships.

5.8.4.1 Johansen Methodology

The Johansen methodology is the dominant technique used to determine whether a system of $I(1)$ variables is cointegrated and if so, to determine the number of cointegrating relationships. Recall that one of the requirements for a set of integrated variables to be cointegrated is that π has reduced rank,

$$\Delta \mathbf{Y}_t = \pi \mathbf{Y}_{t-1} + \pi_1 \Delta \mathbf{Y}_{t-1} + \ldots + \pi_P \Delta \mathbf{Y}_{t-P} \varepsilon_t,$$

and the number of non-zero eigenvalues of π is between 1 and $k - 1$. If the number of non-zero eigenvalues is k, the system is stationary. If no non-zero eigenvalues are present, then the system contains k unit roots, is not cointegrated and it is not possible to define a long-run relationship. The Johansen framework for cointegration analysis uses the magnitude of the eigenvalues of $\hat{\pi}$ to test for cointegration. The Johansen methodology also allows the number of cointegrating relationships to be determined from the data directly, a key feature missing from the Engle-Granger two-step procedure.

The Johansen methodology makes use of two statistics, the trace statistic (λ_{trace}) and the maximum eigenvalue statistic (λ_{max}). Both statistics test functions of the estimated eigenvalues of π but have different null and alternative hypotheses. The trace statistic tests the null that the number of cointegrating relationships is less than or equal to r against an alternative that the number is greater than r.

Define $\hat{\lambda}_i$, $i = 1, 2, \ldots, k$ to be the complex modulus of the eigenvalues of $\hat{\pi}_1$ and let them be ordered such that $\lambda_1 > \lambda_2 > \ldots > \lambda_k$.[18] The trace statistic is defined

$$\lambda_{\text{trace}}(r) = -T \sum_{i=r+1}^{k} \ln\left(1 - \hat{\lambda}_i\right).$$

There are k trace statistics. The trace test is applied sequentially, and the number of cointegrating relationships is determined by proceeding through the test statistics until the null is not rejected. The first trace statistic, $\lambda_{\text{trace}}(0) = -T \sum_{i=1}^{k} \ln(1 - \hat{\lambda}_i)$, tests the null there are no cointegrating relationships (i.e., the system contains k unit roots) against an alternative that the number of cointegrating relationships is one or more. If there are no cointegrating relationships, then the true rank of π is 0, and each of the estimated eigenvalues should be close to zero. The test statistic $\lambda_{\text{trace}}(0) \approx 0$ since every unit root "driver" corresponds to a zero eigenvalue in π. When the series are cointegrated, π has one or more non-zero eigenvalues. If only one eigenvalue is non-zero, so that $\lambda_1 > 0$, then in large samples $\ln\left(1 - \hat{\lambda}_1\right) < 0$ and $\lambda_{\text{trace}}(0) \approx -T(1 - \lambda_1)$, which becomes arbitrarily large as T grows.

Like unit root tests, cointegration tests have nonstandard distributions that depend on the included deterministic terms if any. Software packages return the appropriate critical values for the length of the time-series analyzed and included deterministic regressors if any.

The maximum eigenvalue test examines the null that the number of cointegrating relationships is r against the alternative that the number is $r + 1$. The maximum eigenvalue statistic is defined

$$\lambda_{\text{max}}(r, r+1) = -T \ln\left(1 - \hat{\lambda}_{r+1}\right)$$

Intuitively, if there are $r + 1$ cointegrating relationships, then the $r + 1^{\text{th}}$ ordered eigenvalue should be positive, $\ln\left(1 - \hat{\lambda}_{r+1}\right) < 0$, and the value of $\lambda_{\text{max}}(r, r+1) \approx -T \ln(1 - \lambda_{r+1})$ should be large. On the other hand, if there are only r cointegrating relationships, the $r + 1^{\text{th}}$ eigenvalue is zero, its estimate should be close to zero, and so the statistic should be small. Again, the distribution is nonstandard, but statistical packages provide appropriate critical values for the number of observations and the included deterministic regressors.

The steps to implement the Johansen procedure are:

Step 1: Plot the data series being analyzed and perform univariate unit root testing. A set of variables can only be *cointegrated* if they are all *integrated*. If the series are trending, either linearly or quadratically, remember to include deterministic terms when estimating the VECM.

Step 2: The second stage is lag length selection. Select the lag length using one of the procedures outlined in the VAR lag length selection section (e.g., General-to-Specific or AIC). For example, to use the General-to-Specific approach, first select a maximum lag length L and then, starting with $l = L$, test l lags against $l - 1$ use a likelihood ratio test,

$$LR = (T - l \cdot k^2)(\ln|\Sigma_{l-1}| - \ln|\Sigma_l|) \sim \chi_k^2.$$

Repeat the test by decreasing the number of lags (l) until the LR rejects the null that the smaller model is equivalent to the larger model.

Step 3: Estimate the selected VECM,

---

[18]The complex modulus is defined as $|\lambda_i| = |a + bi| = \sqrt{a^2 + b^2}$.

$$\Delta \mathbf{Y}_t = \pi \mathbf{Y}_{t-1} + \pi_1 \Delta \mathbf{Y}_{t-1} + \ldots + \pi_{P-1} \Delta \mathbf{Y}_{t-P+1} + \varepsilon$$

and determine the rank of π where P is the lag length previously selected. If the levels of the series appear to be trending, then the model in differences should include a constant and

$$\Delta \mathbf{Y}_t = \delta_0 + \pi \mathbf{Y}_{t-1} + \pi_1 \Delta \mathbf{Y}_{t-1} + \ldots + \pi_{P-1} \Delta \mathbf{Y}_{t-P+1} + \varepsilon$$

should be estimated. Using the λ_{trace} and λ_{max} tests, determine the cointegrating rank of the system. It is important to check that the residuals are weakly correlated – so that there are no important omitted variables, the residuals are not excessively heteroskedastic, which affects the size and power of the procedure, and are approximately Gaussian.

Step 4: Analyze the normalized cointegrating vectors to determine whether these conform to implications of finance theory. Hypothesis tests on the cointegrating vector can also be performed to examine whether the long-run relationships conform to a particular theory.

Step 5: The final step of the procedure is to assess the adequacy of the model by plotting and analyzing the residuals. This step should be the final task in the analysis of any time-series data, not just the Johansen methodology. If the residuals do not resemble white noise, the model should be reconsidered. If the residuals are stationary but autocorrelated, more lags may be necessary. If the residuals are $I(1)$, the system may not be cointegrated.

Lag Length Selection

Tests of cointegration using the two test statistic, λ_{trace} and λ_{max}, are sensitive to the lag length. The number of included lags must be sufficient to produce white noise residuals. The lag length is commonly chosen using an IC, and given the trade-off between a model that is too small – which leaves serial correlation in the model residuals – and too large, which produces noisier estimates of parameters but no serial correlation, a loose criterion like the AIC is preferred to a more strict one.

Trends

Nonstationary time series often contain time trends. Like the Augmented Dickey-Fuller test, Johansen's λ_{trace} and λ_{max} tests are both sensitive to the choice of included trends. There are five different configurations of trends in the VECM,

$$\Delta \mathbf{Y}_t = \delta_0 + \delta_1 t + \alpha' (\beta \mathbf{Y}_{t-1} + \gamma_0 + \gamma_1 t) + \pi_1 \Delta \mathbf{Y}_{t-1} + \ldots + \pi_{P-1} \Delta \mathbf{Y}_{t-P+1} + \varepsilon.$$

The five test configurations of the test are:

- no trends, $\delta_0 = \delta_1 = \gamma_0 = \gamma_1 = \mathbf{0}$;

- linear trend in $\mathbf{Y}_t$, $\alpha' \beta \mathbf{Y}_{t-1}$ is mean 0, $\delta_1 = \gamma_0 = \gamma_1 = \mathbf{0}$;

- linear trend in $\mathbf{Y}_t$, non-zero mean $\alpha' \beta \mathbf{Y}_{t-1}$, $\delta_1 = \gamma_1 = \mathbf{0}$;

- quadratic trend in $\mathbf{Y}_t$, non-zero mean $\alpha' \beta \mathbf{Y}_{t-1}$, $\gamma_1 = \mathbf{0}$; and

- quadratic trend in $\mathbf{Y}_t$, linear trend in $\alpha' \beta \mathbf{Y}_{t-1}$, no restrictions on the parameters.

The simplest specification sets all trends to be 0, so that

$$\Delta \mathbf{Y}_t = \alpha' \beta \mathbf{Y}_{t-1} + \pi_1 \Delta \mathbf{Y}_{t-1} + \ldots + \pi_{P-1} \Delta \mathbf{Y}_{t-P+1} + \varepsilon.$$

The specification is only appropriate if the components of $\mathbf{Y}_t$ are not trending. When the component time series of $\mathbf{Y}_t$ have linear time trends, then

$$\Delta \mathbf{Y}_t = \delta_0 + \alpha' (\beta \mathbf{Y}_{t-1} + \gamma_0) + \pi_1 \Delta \mathbf{Y}_{t-1} + \ldots + \pi_{P-1} \Delta \mathbf{Y}_{t-P+1} + \varepsilon$$

allows them to appear in two places. The two intercepts, δ_0 and $\alpha' \gamma_0$ play different roles. δ_0 allows for time trends in the component series since the left-hand-side has been differenced, so that a time-trend in the level becomes an intercept in the difference. γ_0 allows the cointegrating relationship to have a non-zero mean, which is practically important in many applications of cointegration. The model can be estimated assuming $\gamma_0 = \mathbf{0}$ so that

$$\Delta \mathbf{Y}_t = \delta_0 + \alpha' \beta \mathbf{Y}_{t-1} + \pi_1 \Delta \mathbf{Y}_{t-1} + \ldots + \pi_{P-1} \Delta \mathbf{Y}_{t-P+1} + \varepsilon.$$

In this specification, the components are allowed to have unrestricted time trends but the cointegrating relationships are restricted to be mean zero. In practice, this requires that the growth rates of the component time series in $\mathbf{Y}_t$ are the same. The full set of time trends are included in the model, the $\mathbf{Y}_t$ is allowed to have a quadratic time trend (the difference has a linear time trend) and the cointegrating relationship,

$$\beta \mathbf{Y}_{t-1} + \gamma_0 + \gamma_1 t$$

may also have a time trend. The specification with a time trend can be restricted so that $\gamma_1 = \mathbf{0}$ in which case the cointegrating relationships are allowed to have a mean different from $\mathbf{0}$ but not to be trending.

Additional trend components increase the critical values of the λ_{trace} and λ_{max} test statistics, and so, all things equal, it is harder to reject the null. The principle behind selecting deterministic terms in the Johansen's framework is the same as when including deterministic terms in ADF tests – any deterministic that is present in the data must be included, and failing to include a required deterministic term prevents the null from being rejected even in large samples. Similarly, including more deterministic trends than required lowers the power of the test and so makes it more challenging to find cointegration when it is present. Deterministic trends should be eliminated using a general-to-specific search starting with the full set of terms, and eliminating any that are (jointly) insignificant.

5.8.4.2 Example: Consumption Aggregate Wealth

To illustrate cointegration and error correction, three series which have revived the CCAPM in recent years are examined (Lettau and Ludvigson, 2001a; Lettau and Ludvigson, 2001b). These three series are consumption (c), asset prices (a) and labor income (y). The data are made available by Martin Lettau on his web site, and contain quarterly data from 1952:1 until 2017:3. These series are documented to be cointegrated in published papers, and the cointegrating error is related to expected future returns. When $c - \delta_0 - \beta_a a - \beta_y$ is positive, then consumption is above its long-run trend, and so asset returns are expected to be above average. When this error is negative, then c is relatively low compared to asset values and labor income, and so asset values are too high.

| | | Trace Test | |
Null	Alternative	λ_{trace}	Crit. Val.
$r = 0$	$r \geq 1$	19.06	29.80
$r = 1$	$r \geq 2$	8.68	15.49
$r = 2$	$r = 3$	2.03	3.84

| | | Max Test | |
Null	Alternative	λ_{max}	Crit. Val.
$r = 0$	$r = 1$	10.39	21.13
$r = 1$	$r = 2$	6.64	14.26
$r = 2$	$r = 3$	2.03	3.84

Table 5.5: Results of testing using the Johansen methodology to the cay time series.

The Johansen methodology begins by examining the original data for unit roots. The results in Table 5.6 establish that all series have unit roots using ADF tests. The next step tests eigenvalues of π in the VECM

$$\Delta\mathbf{Y}_t = \delta_0 + \pi\left(\mathbf{Y}_{t-1} + \gamma_0\right) + \pi_1\Delta\mathbf{Y}_{t-1} + \pi_2\Delta\mathbf{Y}_{t-2} + \ldots + + \pi_P\Delta\mathbf{Y}_{t-P} + \varepsilon_t.$$

using λ_{trace} and λ_{max} tests. Table 5.5 contains the results of the two tests. These tests are applied sequentially. The first null hypothesis is not rejected for either test, which indicates that the π has rank 0, and so the system contains three distinct unit roots, and so the variables are not cointegrated.[19]

5.8.4.3 A Single Cointegrating Relationship: Engle-Granger Methodology

The Engle-Granger method exploits the defining characteristic of a cointegrated system with a single cointegrating relationship – if the time series are cointegrated, then a linear combination of the series can be constructed that is stationary. If they are not, then any linear combination remains $I(1)$. When there are two variables, the Engle-Granger methodology begins by specifying the cross-section regression

$$Y_t = \beta x_t + \varepsilon_t$$

where $\hat{\beta}$ can be estimated using OLS. It may be necessary to include a constant,

$$Y_t = \delta_0 + \beta x_t + \varepsilon_t$$

or a constant and time trend,

$$Y_t = \delta_0 + \delta_1 t + \beta x_t + \varepsilon_t,$$

if the residuals from the simple cross-sectional regression are not mean 0 or trending. The model residuals, $\hat{\varepsilon}_t$, are constructed from the OLS estimates of the model coefficients and are tested for the

---

[19]The first null not rejected indicates the cointegrating rank of the system. If all null hypotheses are rejected, then the original system appears stationary, and a reanalysis of the $I(1)$ classification of the original data is warranted.

presence of a unit root. If x_t and Y_t are both $I(1)$ and $\hat{\varepsilon}_t$ is $I(0)$, then the series are cointegrated. If the null that $\hat{\varepsilon}_t$ contains a unit root is not rejected, then the two series are no cointegrated since the difference did not eliminate the unit root. The procedure concludes by using $\hat{\varepsilon}_t$ to estimate the VECM to estimate parameters which may be of interest (e.g., the speed of convergence parameters).

Step 1: Begin by analyzing x_t and Y_t in isolation to ensure that they are both integrated, plot the data, and perform ADF tests. Remember, variables can only be *cointegrated* if they are *integrated*.

Step 2: Estimate the long-run relationship by fitting

$$Y_t = \delta_0 + \delta_1 t + \beta x_t + \varepsilon_t,$$

where the two deterministic terms are included only if necessary, using OLS and computing the estimated residuals $\{\hat{\varepsilon}_t\}$. Use an ADF test (or DF-GLS for more power) and test $H_0 : \gamma = 0$ against $H_1 : \gamma < 0$ in the regression

$$\Delta\hat{\varepsilon}_t = \gamma\hat{\varepsilon}_{t-1} + \psi_1\Delta\hat{\varepsilon}_{t-1} + \ldots + \psi_p\Delta\hat{\varepsilon}_{t-P} + \eta_t.$$

Deterministic effects are removed in the cross-sectional regression, and so are not included in the ADF test. If the null is rejected and $\hat{\varepsilon}_t$ is stationary, then x_t and Y_t appear to be cointegrated. Alternatively, if $\hat{\varepsilon}_t$ still contains a unit root, the series are not cointegrated.[20]

Step 3: If a cointegrating relationship is found, specify and estimate the VECM

$$\begin{bmatrix} \Delta x_t \\ \Delta Y_t \end{bmatrix} = \begin{bmatrix} \pi_{01} \\ \pi_{02} \end{bmatrix} + \begin{bmatrix} \alpha_1(Y_{t-1} - \delta_0 - \delta_1 t - \beta x_{t-1}) \\ \alpha_2(Y_{t-1} - \delta_0 - \delta_1 t - \beta x_{t-1}) \end{bmatrix} + \pi_1 \begin{bmatrix} \Delta x_{t-1} \\ \Delta Y_{t-1} \end{bmatrix} + \ldots + \pi_P \begin{bmatrix} \Delta x_{t-P} \\ \Delta Y_{t-P} \end{bmatrix} + \begin{bmatrix} \eta_{1,t} \\ \eta_{2,t} \end{bmatrix}$$

Note that this specification is not linear in its parameters. Both equations have interactions between the α and β parameters and so OLS cannot be used. Engle and Granger noted that the terms involving β can be replaced with $\hat{\varepsilon}_{t-1} = (Y_{t-1} - \hat{\beta}_1 - \hat{\beta}_2 x_{t-1})$,

$$\begin{bmatrix} \Delta x_t \\ \Delta Y_t \end{bmatrix} = \begin{bmatrix} \pi_{01} \\ \pi_{02} \end{bmatrix} + \begin{bmatrix} \alpha_1\hat{\varepsilon}_{t-1} \\ \alpha_2\hat{\varepsilon}_{t-1} \end{bmatrix} + \pi_1 \begin{bmatrix} \Delta x_{t-1} \\ \Delta Y_{t-1} \end{bmatrix} + \ldots + \pi_P \begin{bmatrix} \Delta x_{t-P} \\ \Delta Y_{t-P} \end{bmatrix} + \begin{bmatrix} \eta_{1,t} \\ \eta_{2,t} \end{bmatrix},$$

and so parameters of these specifications can be estimated using OLS. The substitution has no impact on the standard errors of the estimated parameters since the parameters of the cointegrating relationship are super-consistent (i.e., they converge faster than the standard $\sqrt{T}$ rate).

Step 4: The final step is to assess the model adequacy and test hypotheses about α_1 and α_2. Standard diagnostic checks including plotting the residuals and examining the ACF should be used to examine model adequacy. Impulse response functions for the short-run deviations can be examined to assess the effect of a shock on the deviation of the series from the long term trend.

Deterministic Regressors

The cross-sectional regression in the Engle-Granger methodology can be modified to accommodate three configurations of deterministic regressors. The simplest configuration has no deterministic terms so that the regression is

$$Y_t = \beta x_t + \varepsilon_t.$$

---

[20]The distribution of the ADF is different when testing cointegration than when testing for a unit root. Software packages report the correct value which depends on the number of variables in the cointegrating relationship and the deterministic terms if any.

Engle-Granger is only limited finding a single cointegrating relationship, which might exist between k variables, not just 2. In this case, the cross-sectional regression is

$$Y_t = \beta' \mathbf{X}_t + \varepsilon_t$$

where $\mathbf{X}_t$ is the $k-1$ by 1 vector, and the cointegrating vector is $\left[1, -\beta'\right]$. This generalized form can be further extended by altering the deterministic terms in the model. For example, it is common to include an intercept in the cross-sectional regression,

$$Y_t = \delta_0 + \beta' \mathbf{X}_t + \varepsilon_t.$$

This structure allows the long-run relationship between Y_t and $\mathbf{X}_t$ to have a non-zero mean. The intercept should be included except where theory suggests that the cointegrating errors should be zero, e.g., in the relationship between spot and future prices or the long-run relationship between prices of the same security trading in different markets.

The cross-sectional regression can be further extended to include a time trend,

$$Y_t = \delta_0 + \delta_1 t + \beta' \mathbf{X}_t + \varepsilon_t.$$

When the model includes a time-trend, the long-run relationship, Y_t and $\mathbf{X}_t$, is assumed to be trending over time, so that $Y_t - \delta_0 - \delta_1 t - \beta' \mathbf{X}_t$ is a mean-zero $I(0)$ process. This might occur if the growth rates of Y_t and the components $\mathbf{X}_t$ differ. It is much less common to include time-trends in the cointegrating relationship. Best practice is to only include δ_1 if there is some *a priori* reason to believe that the relationship has a time-trend and when $\hat{\delta}_1$ is statistically different from 0 when the cross-sectional regression is estimated. The cross-sectional regression can be compactly expressed as

$$Y_t = \delta' \mathbf{d}_t + \beta' \mathbf{X}_t + \varepsilon_t$$

where $\mathbf{d}_t$ is the vector of included deterministic regressors, i.e., on of $[]$ (nothing), $[1]$, or $[1,t]$.

Dynamic OLS

The parameter estimators of the cointegrating vector estimated using a cross-sectional regression is not normally distributed in large samples. It is also not efficient since the $I(1)$ variables might have short-run dynamics. Dynamic OLS, a simple modification of the Engle-Granger regression, addresses both of these. It adds lags and *leads* of the differences on the right-hand-side variables to the cross-sectional regression. These extra terms effectively remove the short term dynamics in the right-hand-side variables. In a bivariate cointegrated relationship, the Dynamic OLS regression is

$$Y_t = \delta' \mathbf{d}_t + \beta_1 x_t + \sum_{i=-P}^{P} \gamma_i \Delta x_{t-i} + \varepsilon_t$$

where $\mathbf{d}_t$ is a vector of deterministic terms in the model. This regression is estimated using OLS, and the estimated cointegrating relationship is $Y_t - \hat{\delta}' \mathbf{d}_t - \hat{\beta}_1 x_t$. If there are more than 1-right-hand-side variables, then the regression is

$$Y_t = \delta' \mathbf{d}_t + \beta' \mathbf{X}_t + \sum_{i=-P}^{P} \gamma' \Delta \mathbf{X}_{t-i} + \varepsilon_t$$

Unit Root Tests			
Series	T-stat	P-val	ADF Lags
c	-1.198	0.674	5
a	-0.205	0.938	3
y	-2.302	0.171	0
$\hat{\varepsilon}_t^c$	-2.706	0.379	1
$\hat{\varepsilon}_t^a$	-2.573	0.447	0
$\hat{\varepsilon}_t^y$	-2.679	0.393	1

Table 5.6: The top three lines contain the results of ADF tests for unit roots in the three components of cay : Consumption, Asset Prices and Aggregate Wealth. The final lines contain the results of unit root tests on the estimated residuals from the cross-sectional regressions. The variable in the superscript is the dependent variable in the Engle-Granger regression. The lags column reports the number of lags used in the ADF procedure, which is automatically selected using the AIC.

where β, γ_i and $\mathbf{X}_t$ are $k-1$ by 1 vectors. The estimators of the cointegrating vector are asymptotically normally distributed, although the parameter covariance must be estimated using a long-run covariance estimator that accounts for dependence, e.g., Newey-West (see Section 5.9.2). The number of leads and lags to include in the model is can be selected using an information criterion. In application in macrofinance, it is often chosen to capture 1 year of data, so either 4 (quarterly) or 12 (monthly).

5.8.4.4 Cointegration in Consumption, Asset Prices and Income

The Engle-Granger procedure begins by performing unit root tests on the individual series and examining the data. Table 5.6 and contain the results from ADF tests and Figure 5.6 plots the detrended series. The null of a unit root is not rejected in any of the three series, and all have time-detrended errors which appear to be nonstationary.

The next step is to specify the cointegrating regression

$$c_t = \delta_0 + \beta_a a_t + \beta_y Y_t + \varepsilon_t$$

and to estimate the long-run relationship using OLS. The estimated cointegrating vector from is $[1 - 0.249 - 0.785]$, and corresponds to a long-run relationship of $\hat{\varepsilon}_t = c_t + .643 - 0.249a_t - 0.785Y_t$. Finally, the residuals are tested for the presence of a unit root. The results of this test are labeled $\hat{\varepsilon}_t^c$ in Table 5.6 and indicate that the null is not rejected, and so the three series are not cointegrated. The Engle-Granger methodology agrees with the Johansen methodology that it is not possible to eliminate the unit roots from the three series using a single linear combination. It is also possible to normalize the coefficients on a or y by using these are the dependent variable. The final two lines in Table 5.6 contain results for these specifications. The results for the alternative agree with the finding for c, and the series do not appear to be cointegrated. The middle panel of Figure 5.6 plot the three residual series where each of the variables is used as the dependent. The residuals constructed from the regression when a or y are the dependent are multiplied by -1 so that the sign on c is always positive, and all three series are normalized to have unit variance (for comparability). The three residual series are very

Comparing Engle-Granger and Dynamic OLS

	Dependent Variable					
	c		a		y	
δ_0	-0.643	-0.640 (-6.896)	1.917	1.874 (7.784)	0.702	0.713 (5.312)
β_c			2.284	2.385 (6.994)	1.163	1.180 (18.521)
β_a	0.249	0.260 (6.187)			-0.214	-0.229 (-3.790)
β_y	0.785	0.773 (17.339)	-1.322	-1.421 (-4.024)		

Table 5.7: Each column reports estimates of the cointegrating relationship where the dependent variable varies across the three series. The parameter estimators in Engle-Granger regressions are not asymptotically normally distributed, and so t-stats are not reported. The t-stats reported for the estimates produces using Dynamic OLS are computed using the Newey-West covariance estimator with 14 lags.

similar which indicates that the choice of the dependent variable has little impact on the estimates of the cointegrating relationship.

The VECM uses the residuals estimated using the cross-sectional regression, $\hat{\varepsilon}_t = c_t - \hat{\delta}_0 - \hat{\beta}_a a_t - \hat{\beta}_y Y_t$.

$$
\begin{bmatrix} \Delta c_t \\ \Delta a_t \\ \Delta y_t \end{bmatrix} = \begin{bmatrix} 0.003 \\ (0.000) \\ 0.004 \\ (0.014) \\ 0.003 \\ (0.000) \end{bmatrix} + \begin{bmatrix} -0.000 \\ (0.281) \\ 0.002 \\ (0.037) \\ 0.000 \\ (0.515) \end{bmatrix} \hat{\varepsilon}_{t-1} + \begin{bmatrix} 0.192 & 0.102 & 0.147 \\ (0.005) & (0.000) & (0.004) \\ 0.282 & 0.220 & -0.149 \\ (0.116) & (0.006) & (0.414) \\ 0.369 & 0.061 & -0.139 \\ (0.000) & (0.088) & (0.140) \end{bmatrix} \begin{bmatrix} \Delta c_{t-1} \\ \Delta a_{t-1} \\ \Delta y_{t-1} \end{bmatrix} + \eta_t
$$

The coefficients on the lagged residual measure the speed of adjustment. The estimates are all close to 0 indicating that deviations from the equilibrium are highly persistent. Two of the speed of adjustment coefficients are not statistically different from zero, which indicates that three series are not well described as a cointegrated system. The lag length in the VECM is selected by minimizing the HQIC using up to 4 lags of the quarterly data.

Table 5.7 contains estimates of the parameters from the Engle-Granger cross-sectional regressions and the Dynamic OLS regressions. The DOLS estimates are asymptotically normal (if the series are cointegrated) and so standard errors, computed using the Newey-West covariance estimator, are reported for the coefficients. The bottom panel of Figure 5.6 plot the residual from the two estimators when c is the dependent variable. The leads and lags have little effect on the estimated cointegration vector, and so the two series are very similar.

5.8.5   Spurious Regression and Balance

When a regression is estimated using two related $I(1)$ variables, the cointegrating relationship dominates and the regression coefficients can be directly interpreted as the cointegrating vectors. However,

Analysis of *cay*

Residuals
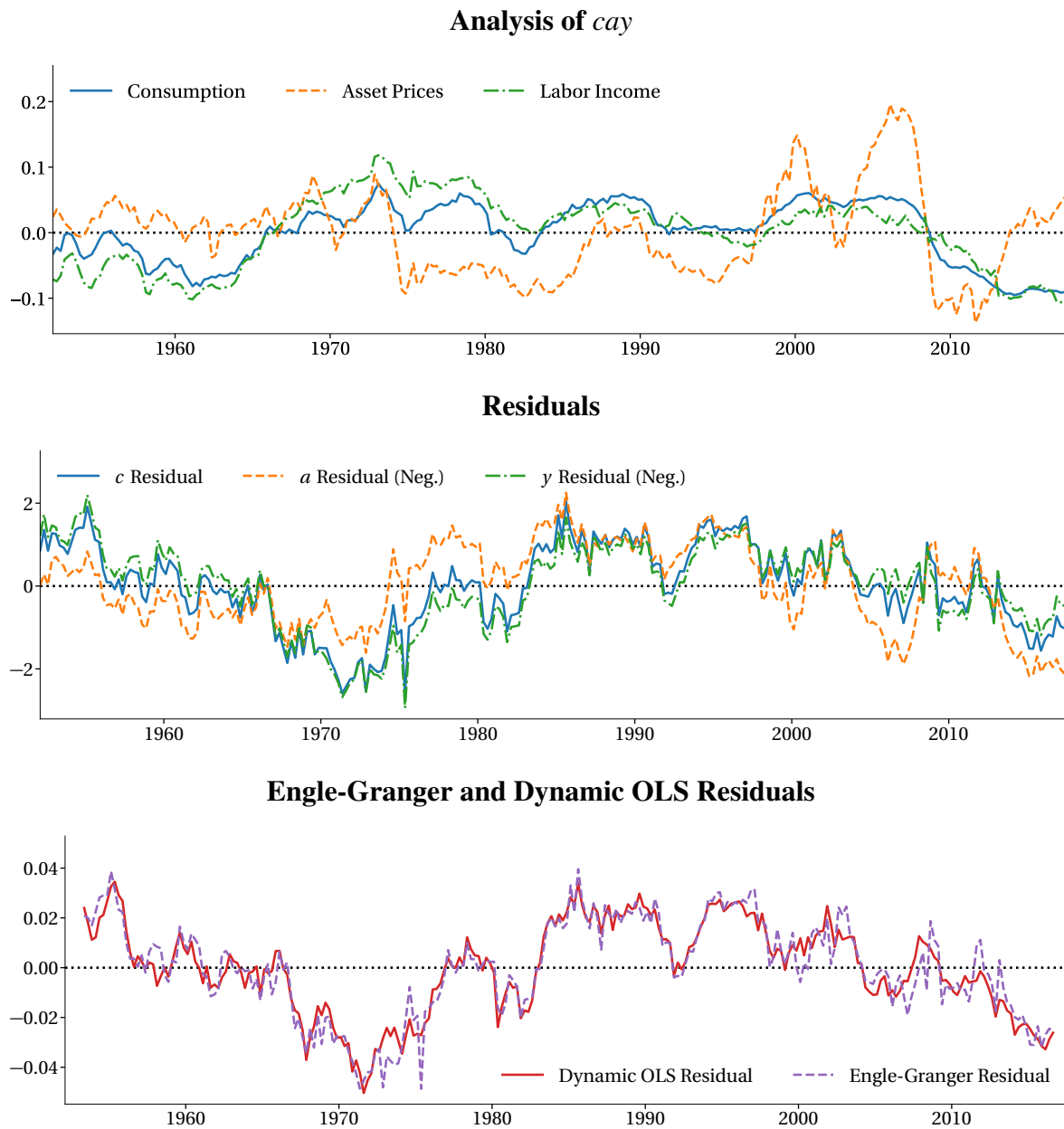

Engle-Granger and Dynamic OLS Residuals

Figure 5.6: The top panel contains plots of detrended residuals from regressions of consumption, asset prices and labor income on a linear time trend. The middle panel contains a plot of residuals from the three specifications of the Engle-Granger regression where each of the three series is used as the dependent variable. The residuals are multiplied by -1 when a or y is the dependent variable so they the sign on c is always positive. The residuals are all normalized to have unit variance. The bottom panel plots the residuals computed using the Dynamic OLS estimates of the cointegrating relationship when c is the dependent variable and 4 leads and lags are used.

when a model is estimated on two unrelated $I(1)$ variables, the regression estimator is no longer consistent. For example, let x_t and Y_t be independent random walk processes.

$$x_t = x_{t-1} + \eta_t$$

and

$$Y_t = Y_{t-1} + \nu_t$$

In the regression

$$x_t = \beta Y_t + \varepsilon_t$$

$\hat{\beta}$ is not consistent for 0 despite the independence of x_t and Y_t.

Models that include independent $I(1)$ processes are known as **spurious regressions**. When the regressions are spurious, the estimated $\hat{\beta}$ can take any value and typically have t-stats that indicate significance at conventional levels. The solution to this problems is simple: whenever regressing one $I(1)$ variable on another, always check to be sure that the regression residuals are $I(0)$ and not $I(1)$ – in other words, verify that the series are cointegrated. If the series are not cointegrated, it is not possible to estimate a meaningful long-run relationship between the two (or more) $I(1)$ random variables. Nonstationary time series that are not cointegrated can be differenced to be $I(0)$ and then modeled as a stationary VAR.

Balance is an important concept when data which contain both stationary and integrated data. An equation is balanced if all variables have the same order of integration. The usual case occurs when a stationary variable ($I(0)$) is related to one or more other stationary variables. It is illustrative to consider the four combinations:

- $I(0)$ on $I(0)$: The usual case. Standard asymptotic arguments apply. See section 5.9 for more issues in cross-section regression using time-series data.

- $I(1)$ on $I(0)$: This regression is unbalanced. An $I(0)$ variable can never explain the long-run variation in an $I(1)$ variable. The usual solution is to difference the $I(1)$ and then examine whether the short-run dynamics in the differenced $I(1)$, which are $I(0)$, can be explained by the $I(0)$.

- $I(1)$ on $I(1)$: One of two outcomes: cointegration or spurious regression.

- $I(0)$ on $I(1)$: This regression is unbalanced. An $I(1)$ variable can never explain the variation in an $I(0)$ variable, and unbalanced regressions are not useful tools for explaining economic phenomena. Unlike spurious regressions, the t-stat still has a standard asymptotic distribution although caution is needed since the CLT does not, in empirically relevant samples sizes, provide an accurate approximation to the finite sample distribution. Poor finite-sample approximations are common in applications where a stationary variable, e.g., returns on the market, is regressed on a highly persistent predictor (such as the default premium, dividend yield or price-to-earnings ratio).

5.9 Cross-sectional Regression with Time-series Data

Cross-sectional regressions are commonly estimated using data that occur sequentially, e.g., the CAP-M and related models. Chapter 3 used n to index the observations to indicate that the data are not ordered,

$$Y_n = \beta_1 x_{n1} + \beta_2 x_{n2} + \ldots + \beta_k x_{nk} + \varepsilon_n. \tag{5.39}$$

Here the observation index is replaced with t to indicate that ordered time-series data are used in the regression,

$$Y_t = \beta_1 x_{t1} + \beta_2 x_{t2} + \ldots + \beta_k x_{tk} + \varepsilon_t. \tag{5.40}$$

Five assumptions are used to establish the asymptotic distribution of the parameter estimated. Here these assumptions are restated using time-series indices.

Assumption 5.1 (Linearity). *The model specification is linear in* $\mathbf{X}_t$, $Y_t = \mathbf{X}_t \beta + \varepsilon_t$.

Assumption 5.2 (Stationary Ergodicity). $\{(\mathbf{X}_t, \varepsilon_t)\}$ *is a strictly stationary and ergodic sequence.*

Assumption 5.3 (Rank). $\mathrm{E}[\mathbf{X}_t' \mathbf{X}_t] = \Sigma_{\mathbf{XX}}$ *is non-singular and finite.*

Assumption 5.4 (Martingale Difference). $\{\mathbf{X}_t' \varepsilon_t, \mathcal{F}_{t-1}\}$ *is a martingale difference sequence,* $\mathrm{E}\left[\left(x_{j,t} \varepsilon_t \right)^2 \right] < \infty$ $j = 1, 2, \ldots, k$, $t = 1, 2 \ldots$ *and* $\mathbf{S} = \mathrm{V}[T^{-\frac{1}{2}} \mathbf{X}' \varepsilon]$ *is finite and non singular.*

Assumption 5.5 (Moment Existence). $\mathrm{E}[x_{j,t}^4] < \infty$, $j = 1, 2, \ldots, k$, $t = 1, 2, \ldots$ *and* $\mathrm{E}[\varepsilon_t^2] = \sigma^2 < \infty$, $t = 1, 2, \ldots$.

Assumption 3.9 may be violated when estimating cross-sectional models using time series data. When this assumption is violated, the scores from the linear regression, $\mathbf{X}_t' \varepsilon_t$ are a not martingale difference with respect to the time $t - 1$ information set, $\mathcal{F}_{t-1}$. The autocorrelation in the scores occurs when the errors from the model, ε_t, have a persistent component that is not explained by the regressors. The MDS assumption featured prominently in two theorems: the asymptotic distribution of $\hat{\beta}$ and the estimation of the covariance of the parameters.

Theorem 5.5. *Under assumptions 3.1 and 3.7 - 3.9*

$$\sqrt{T}(\hat{\beta}_T - \beta) \xrightarrow{d} N(0, \Sigma_{\mathbf{XX}}^{-1} \mathbf{S} \Sigma_{\mathbf{XX}}^{-1}) \tag{5.41}$$

where $\Sigma_{\mathbf{XX}} = \mathrm{E}[\mathbf{X}_t' \mathbf{X}_t]$ *and* $\mathbf{S} = \mathrm{V}[T^{-1/2} \mathbf{X}' \varepsilon]$

Theorem 5.6. *Under assumptions 3.1 and 3.7 - 3.10,*

$$\hat{\Sigma}_{\mathbf{XX}} = T^{-1} \mathbf{X}' \mathbf{X} \xrightarrow{p} \Sigma_{\mathbf{XX}}$$

$$\hat{\mathbf{S}} = T^{-1} \sum_{n=1}^{T} e_t^2 \mathbf{X}_t' \mathbf{X}_t \xrightarrow{p} \mathbf{S}$$

$$= T^{-1} \left(\mathbf{X}' \hat{\mathbf{E}} \mathbf{X} \right)$$

and

$$\hat{\Sigma}_{\mathbf{XX}}^{-1} \hat{\mathbf{S}} \hat{\Sigma}_{\mathbf{XX}}^{-1} \xrightarrow{p} \Sigma_{\mathbf{XX}}^{-1} \mathbf{S} \Sigma_{\mathbf{XX}}^{-1}$$

where $\hat{\mathbf{E}} = \mathrm{diag}(\hat{\varepsilon}_1^2, \ldots, \hat{\varepsilon}_T^2)$ *is a matrix with the squared estimated residuals along the diagonal.*

When the MDS assumption does now hold, the asymptotic covariance takes a different form that reflects the persistence in the data, and so an alternative estimator is required to estimate the covariance of $\hat{\beta}$. The new estimator is an extended version of White's covariance estimator that accounts for the predictability of the scores ($\mathbf{X}'_t \varepsilon_t$). The correlation in the scores alters the amount of "unique" information available to estimate the parameters. The standard covariance estimator assumes that the scores are uncorrelated with their past and so each contributes its full share to the precision to $\hat{\beta}$. When the scores are autocorrelated, only the unpredictable component of the score is informative about the value of the regression coefficient, and the covariance estimator must account for this change in the available information. Heteroskedasticity Autocorrelation Consistent (HAC) covariance estimators are consistent even in the presence of score autocorrelation.

5.9.1 Estimating the mean with time-series errors

To understand why a HAC estimator is needed, consider estimating the mean in two different setups. In the first, the shock, $\{\varepsilon_t\}$, is assumed to be a white noise process with variance σ^2. In the second, the shock follows an MA(1) process.

5.9.1.1 White Noise Errors

Suppose the data generating process for Y_t is,

$$Y_t = \mu + \varepsilon_t$$

where $\{\varepsilon_t\}$ is a white noise process. It is simple to show that

$$E[Y_t] = \mu \text{ and } V[Y_t] = \sigma^2$$

since the error is a white noise process. Define the sample mean estimator in the usual way,

$$\hat{\mu} = T^{-1} \sum_{t=1}^{T} Y_t$$

The sample mean is unbiased,

$$
\begin{aligned}
E[\hat{\mu}] &= E\left[T^{-1} \sum_{t=1}^{T} Y_t \right] \\
&= T^{-1} \sum_{t=1}^{T} E[Y_t] \\
&= T^{-1} \sum_{t=1}^{T} \mu \\
&= \mu.
\end{aligned}
$$

The variance of the mean estimator exploits the white noise property which ensures $E[\varepsilon_i \varepsilon_j]=0$ whenever $i \neq j$.

$$V[\hat{\mu}] = E[(T^{-1}\sum_{t=1}^{T} Y_t - \mu)^2]$$

$$= E[(T^{-1}\sum_{t=1}^{T} \varepsilon_t)^2]$$

$$= E[T^{-2}(\sum_{t=1}^{T} \varepsilon_t^2 + \sum_{r=1}^{T} \sum_{s=1,r\neq s}^{T} \varepsilon_r \varepsilon_s)]$$

$$= T^{-2}\sum_{t=1}^{T} E[\varepsilon_t^2] + T^{-2}\sum_{r=1}^{T} \sum_{s=1,r\neq s}^{T} E[\varepsilon_r \varepsilon_s]$$

$$= T^{-2}\sum_{t=1}^{T} \sigma^2 + T^{-2}\sum_{r=1}^{T} \sum_{s=1,r\neq s}^{T} 0$$

$$= T^{-2}T\sigma^2$$

$$= \sigma^2/T,$$

and so, $V[\hat{\mu}] = \sigma^2/T$, the standard result.

5.9.1.2 MA(1) errors

Suppose the model is altered so that the error process $(\{\eta_t\})$ is a mean zero MA(1) constructed from white noise shocks $(\{\varepsilon_t\})$,

$$\eta_t = \theta \varepsilon_{t-1} + \varepsilon_t.$$

The properties of the error are easily derived using the results in Chapter 4. The mean is 0,

$$E[\eta_t] = E[\theta \varepsilon_{t-1} + \varepsilon_t] = \theta E[\varepsilon_{t-1}] + E[\varepsilon_t] = \theta 0 + 0 = 0,$$

and the variance depends on the MA parameter,

$$V[\eta_t] = E[(\theta \varepsilon_{t-1} + \varepsilon_t)^2]$$

$$= E[\theta^2 \varepsilon_{t-1}^2 + 2\varepsilon_t \varepsilon_{t-1} + \varepsilon_t^2]$$

$$= E[\theta^2 \varepsilon_{t-1}^2] + 2E[\varepsilon_t \varepsilon_{t-1}] + E[\varepsilon_t^2]$$

$$= \theta^2 \sigma^2 + 2 \cdot 0 + \sigma^2$$

$$= \sigma^2(1 + \theta^2).$$

The DGP for Y_t is

$$Y_t = \mu + \eta_t,$$

and so the mean and variance of Y_t are

$$E[Y_t] = \mu \text{ and } V[Y_t] = V[\eta_t] = \sigma^2(1 + \theta^2).$$

The sample mean estimator remains unbiased,

$$\hat{\mu} = T^{-1}\sum_{t=1}^{T} Y_t$$

$$\begin{aligned}
\mathrm{E}[\hat{\mu}] &= \mathrm{E}\left[T^{-1}\sum_{t=1}^{T} Y_t\right] \\
&= T^{-1}\sum_{t=1}^{T}\mathrm{E}[Y_t] \\
&= T^{-1}\sum_{t=1}^{T}\mu \\
&= \mu.
\end{aligned}$$

The variance of the mean estimator, however, is different, since η_t is autocorrelated, and so $\mathrm{E}[\eta_t\eta_{t-1}] \neq 0$.

$$\begin{aligned}
\mathrm{V}[\hat{\mu}] &= \mathrm{E}\left[(T^{-1}\sum_{t=1}^{T} Y_t - \mu)^2\right] \\
&= \mathrm{E}\left[(T^{-1}\sum_{t=1}^{T} \eta_t)^2\right] \\
&= \mathrm{E}\left[T^{-2}(\sum_{t=1}^{T}\eta_t^2 + 2\sum_{t=1}^{T-1}\eta_t\eta_{t+1} + 2\sum_{t=1}^{T-2}\eta_t\eta_{t+2} + \ldots + 2\sum_{t=1}^{2}\eta_t\eta_{t+T-2} + 2\sum_{t=1}^{1}\eta_t\eta_{t+T-1})\right] \\
&= T^{-2}\sum_{t=1}^{T}\mathrm{E}[\eta_t^2] + 2T^{-2}\sum_{t=1}^{T-1}\mathrm{E}[\eta_t\eta_{t+1}] + 2T^{-2}\sum_{t=1}^{T-2}\mathrm{E}[\eta_t\eta_{t+2}] + \ldots + \\
&\qquad 2T^{-2}\sum_{t=1}^{2}\mathrm{E}[\eta_t\eta_{t+T-2}] + 2T^{-2}\sum_{t=1}^{1}\mathrm{E}[\eta_t\eta_{t+T-1}] \\
&= T^{-2}\sum_{t=1}^{T}\gamma_0 + 2T^{-2}\sum_{t=1}^{T-1}\gamma_1 + 2T^{-2}\sum_{t=1}^{T-2}\gamma_2 + \ldots + 2T^{-2}\sum_{t=1}^{2}\gamma_{T-2} + 2T^{-2}\sum_{t=1}^{1}\gamma_{T-1}
\end{aligned}$$

where $\gamma_0 = \mathrm{E}[\eta_t^2] = \mathrm{V}[\eta_t]$ and $\gamma_s = \mathrm{E}[\eta_t\eta_{t-s}]$. Only γ_0 and γ_1 are non-zero when the error follows an MA(1) process. $\gamma_0 = \mathrm{V}[\eta_t] = \sigma^2\left(1+\theta^2\right)$ and

$$
\begin{aligned}
\gamma_1 &= \mathrm{E}[\eta_t \eta_{t-1}] \\
&= \mathrm{E}[(\theta\varepsilon_{t-1} + \varepsilon_t)(\theta\varepsilon_{t-2} + \varepsilon_{t-1})] \\
&= \mathrm{E}[\theta^2 \varepsilon_{t-1}\varepsilon_{t-2} + \theta\varepsilon_{t-1}^2 + \theta\varepsilon_t\varepsilon_{t-2} + \varepsilon_t\varepsilon_{t-1}] \\
&= \theta^2 \mathrm{E}[\varepsilon_{t-1}\varepsilon_{t-2}] + \theta\mathrm{E}[\varepsilon_{t-1}^2] + \theta\mathrm{E}[\varepsilon_t\varepsilon_{t-2}] + \mathrm{E}[\varepsilon_t\varepsilon_{t-1}] \\
&= \theta^2 0 + \theta\sigma^2 + \theta 0 + 0 \\
&= \theta\sigma^2.
\end{aligned}
$$

The remaining autocovariance are all 0 since $\gamma_s = 0$, $s > Q$ in a MA(Q). Returning to the variance of $\hat{\mu}$,

$$
\begin{aligned}
\mathrm{V}[\hat{\mu}] &= T^{-2} \sum_{t=1}^{T} \gamma_0 + 2T^{-2} \sum_{t=1}^{T-1} \gamma_1 \tag{5.42} \\
&= T^{-2} T \gamma_0 + 2T^{-2}(T-1)\gamma_1 \\
&\approx \frac{\gamma_0 + 2\gamma_1}{T}.
\end{aligned}
$$

When the errors are autocorrelated, the usual mean estimator has a different variance that reflects the dependence in the errors. Importantly, the usual estimator variance is no longer correct and $\mathrm{V}[\hat{\mu}] \neq \gamma_0/T$.

This simple illustration captures the key idea that underlies the Newey-West variance estimator,

$$
\hat{\sigma}_{NW}^2 = \hat{\gamma}_0 + 2 \sum_{l=1}^{L} \left(1 - \frac{l}{L+1}\right)\hat{\gamma}_l.
$$

When $L = 1$, the only weight is $1 - 1/2 = 1/2$ and $\hat{\sigma}_{NW}^2 = \hat{\gamma}_0 + \hat{\gamma}_1$, which is different from the variance in the MA(1) error example. However as L increases, the weight on γ_1 converges to 1 since $\lim_{L\to\infty} 1 - \frac{1}{L+1} = 1$. The Newey-West variance estimator asymptotically includes all of the autocovariance in the variance, $\gamma_0 + 2\gamma_1$, and when L grows large,

$$
\hat{\sigma}_{NW}^2 \to \gamma_0 + 2\gamma_1.
$$

The variance of the estimated mean can be consistently estimated using σ_{NW}^2 as

$$
\mathrm{V}[\hat{\mu}] = \frac{\gamma_0 + 2\gamma_1}{T} \approx \frac{\sigma_{NW}^2}{T}.
$$

As a general principle, the variance of the sum is the sum of the variances only true when the errors are uncorrelated. HAC covariance estimators account for time-series dependence and lead to correct inference as long as L grows with the sample size.[21]

It is tempting to estimate eq. (5.42) using the natural estimator $\hat{\sigma}_{HAC}^2 = \hat{\gamma}_0 + 2\hat{\gamma}_1/T$. This estimator is not guaranteed to be positive in finite samples, an in general unweighted estimators of the form

---

[21] Allowing L to grow at the rate $T^{1/3}$ is optimal in a certain sense related to testing.

$\hat{\sigma}_{HAC}^2 = \hat{\gamma}_0 + 2\hat{\gamma}_1 + 2\hat{\gamma}_2 + \ldots + 2\hat{\gamma}_L$, may be negative. The Newey-West variance estimator, $\hat{\sigma}_{NW}^2$, is guaranteed to be positive for any L. The weights that scale the autocovariances, $\left(1 - \frac{l}{L+1}\right)$, alter the estimator and ensure that the estimate is positive.

5.9.2   Estimating the variance of $\hat{\beta}$ when the errors are autocorrelated

There are two solutions to modeling cross-sectional data that have autocorrelated errors. The direct method is to alter the cross-sectional model to capture the time-series variation by including both contemporaneous effects of $\mathbf{X}_t$ as well as lagged values of Y_t (and possibly lags of $\mathbf{X}_t$). This approach needs to include sufficient lags so that the errors are white noise. If the dependence is fully modeled, then White's heteroskedasticity (but not autocorrelation) consistent covariance estimator is consistent, and there is no need for a more complex covariance estimator.

The second approach modifies the covariance estimator to account for the dependence in the data. The key insight in White's estimator of $\mathbf{S}$,

$$\hat{\mathbf{S}} = T^{-1} \sum_{t=1}^{T} e_t^2 \mathbf{X}_t' \mathbf{X}_t,$$

is that this form explicitly captures the dependence between the e_t^2 and $\mathbf{X}_t' \mathbf{X}_t$. Heteroskedasticity Autocorrelation Consistent estimators work similarly by capturing both the dependence between the e_t^2 and $\mathbf{X}_t' \mathbf{X}_t$ (heteroskedasticity) and the dependence between the $\mathbf{X}_t e_t$ and $\mathbf{X}_{t-j} e_{t-j}$ (autocorrelation). HAC estimators of the score covariance in linear regressions use the same structure, and

$$\hat{\mathbf{S}}_{HAC} = T^{-1} \left(\sum_{t=1}^{T} e_t^2 \mathbf{X}_t' \mathbf{X}_t + \sum_{l=1}^{L} w_l \left(\sum_{s=l+1}^{T} e_s e_{s-l} \mathbf{X}_s' \mathbf{X}_{s-l} + \sum_{q=l+1}^{T} e_{q-l} e_q \mathbf{X}_{q-l}' \mathbf{X}_q \right) \right) \qquad (5.43)$$

$$= \hat{\Gamma}_0 + \sum_{l=1}^{L} w_l \left(\hat{\Gamma}_l + \hat{\Gamma}_{-l} \right)$$

$$= \hat{\Gamma}_0 + \sum_{l=1}^{L} w_l \left(\hat{\Gamma}_l + \hat{\Gamma}_l' \right)$$

where $\{w_l\}$ are a set of weights. The Newey-West estimator uses $w_l = 1 - \frac{l}{L+1}$ and is always positive semi-definite. Other estimators alter the weights and have different finite-sample properties.

5.A   Cointegration in a trivariate VAR

This section details how to:

- determine whether a trivariate VAR is cointegrated;

- determine the number of cointegrating vectors in a cointegrated system; and

- decompose the π matrix into α, the adjustment coefficient, and β, the cointegrating vectors.

5.A.1 Stationary VAR

Consider the VAR(1),

$$
\begin{bmatrix} x_t \\ Y_t \\ Z_t \end{bmatrix} = \begin{bmatrix} .9 & -.4 & .2 \\ .2 & .8 & -.3 \\ .5 & .2 & .1 \end{bmatrix} \begin{bmatrix} x_{t-1} \\ Y_{t-1} \\ Z_{t-1} \end{bmatrix} + \begin{bmatrix} \varepsilon_{1,t} \\ \varepsilon_{2,t} \\ \varepsilon_{3,t} \end{bmatrix}.
$$

The eigenvalues of the parameter matrix determine the stationarity properties of this VAR process. If the eigenvalues are all less than one in modulus, then the VAR(1) is stationary. This is the case here, and the eigenvalues are 0.97, 0.62, and 0.2. An alternative method is to transform the model into an VECM

$$
\begin{bmatrix} \Delta x_t \\ \Delta Y_t \\ \Delta Z_t \end{bmatrix} = \left(\begin{bmatrix} .9 & -.4 & .2 \\ .2 & .8 & -.3 \\ .5 & .2 & .1 \end{bmatrix} - \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \right) \begin{bmatrix} x_{t-1} \\ Y_{t-1} \\ Z_{t-1} \end{bmatrix} + \begin{bmatrix} \varepsilon_{1,t} \\ \varepsilon_{2,t} \\ \varepsilon_{3,t} \end{bmatrix}
$$

$$
\begin{bmatrix} \Delta x_t \\ \Delta Y_t \\ \Delta Z_t \end{bmatrix} = \begin{bmatrix} -.1 & -.4 & .2 \\ .2 & -.2 & -.3 \\ .5 & .2 & -.9 \end{bmatrix} \begin{bmatrix} x_{t-1} \\ Y_{t-1} \\ Z_{t-1} \end{bmatrix} + \begin{bmatrix} \varepsilon_{1,t} \\ \varepsilon_{2,t} \\ \varepsilon_{3,t} \end{bmatrix}
$$

$$
\Delta \mathbf{w}_t = \pi \mathbf{w}_t + \varepsilon_t
$$

where $\mathbf{w}_t$ is a vector composed of x_t, Y_t and Z_t. The rank of the parameter matrix π can be determined by transforming it into row-echelon form.

$$
\begin{bmatrix} -0.1 & -0.4 & 0.2 \\ 0.2 & -0.2 & -0.3 \\ 0.5 & 0.2 & -0.9 \end{bmatrix} \Rightarrow \begin{bmatrix} 1 & 4 & -2 \\ 0.2 & -0.2 & -0.3 \\ 0.5 & 0.2 & -0.9 \end{bmatrix} \Rightarrow \begin{bmatrix} 1 & 4 & -2 \\ 0 & -1 & 0.1 \\ 0 & -1.8 & 0.1 \end{bmatrix} \Rightarrow \begin{bmatrix} 1 & 4 & -2 \\ 0 & 1 & -0.1 \\ 0 & -1.8 & 0.1 \end{bmatrix}
$$

$$
\Rightarrow \begin{bmatrix} 1 & 0 & -1 \\ 0 & 1 & -0.1 \\ 0 & 0 & -0.08 \end{bmatrix} \Rightarrow \begin{bmatrix} 1 & 0 & -1 \\ 0 & 1 & -0.1 \\ 0 & 0 & 1 \end{bmatrix} \Rightarrow \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}
$$

Since the π matrix is full rank, the system must be stationary. This method is equivalent to computing the eigenvalues of the parameter matrix in the VAR.

5.A.2 Independent Unit Roots

Consider the simple VAR

$$
\begin{bmatrix} x_t \\ Y_t \\ Z_t \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x_{t-1} \\ Y_{t-1} \\ Z_{t-1} \end{bmatrix} + \begin{bmatrix} \varepsilon_{1,t} \\ \varepsilon_{2,t} \\ \varepsilon_{3,t} \end{bmatrix}.
$$

The eigenvalues of the coefficient matrix are all 1 and the VECM is

$$
\begin{bmatrix} \Delta x_t \\ \Delta Y_t \\ \Delta Z_t \end{bmatrix} = \left(\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} - \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \right) \begin{bmatrix} x_{t-1} \\ Y_{t-1} \\ Z_{t-1} \end{bmatrix} + \begin{bmatrix} \varepsilon_{1,t} \\ \varepsilon_{2,t} \\ \varepsilon_{3,t} \end{bmatrix}
$$

$$
\begin{bmatrix} \Delta x_t \\ \Delta Y_t \\ \Delta Z_t \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} x_{t-1} \\ Y_{t-1} \\ Z_{t-1} \end{bmatrix} + \begin{bmatrix} \varepsilon_{1,t} \\ \varepsilon_{2,t} \\ \varepsilon_{3,t} \end{bmatrix}
$$

and the rank of π is clearly 0, so this system contains three independent unit roots. This structure also applies to higher order nonstationary VAR models that contain independent unit root processes – the coefficient matrix in the VECM is always rank 0 when the system contains as many distinct unit roots as variables.

5.A.3   Cointegrated with one cointegrating relationship

Consider the VAR(1),

$$
\begin{bmatrix} x_t \\ Y_t \\ Z_t \end{bmatrix} = \begin{bmatrix} 0.8 & 0.1 & 0.1 \\ -0.16 & 1.08 & 0.08 \\ 0.36 & -0.18 & 0.82 \end{bmatrix} \begin{bmatrix} x_{t-1} \\ Y_{t-1} \\ Z_{t-1} \end{bmatrix} + \begin{bmatrix} \varepsilon_{1,t} \\ \varepsilon_{2,t} \\ \varepsilon_{3,t} \end{bmatrix}.
$$

The eigenvalues of the parameter matrix are 1, 1 and .7. The VECM form of this model is

$$
\begin{bmatrix} \Delta x_t \\ \Delta Y_t \\ \Delta Z_t \end{bmatrix} = \left(\begin{bmatrix} 0.8 & 0.1 & 0.1 \\ -0.16 & 1.08 & 0.08 \\ 0.36 & -0.18 & 0.82 \end{bmatrix} - \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \right) \begin{bmatrix} x_{t-1} \\ Y_{t-1} \\ Z_{t-1} \end{bmatrix} + \begin{bmatrix} \varepsilon_{1,t} \\ \varepsilon_{2,t} \\ \varepsilon_{3,t} \end{bmatrix}
$$

$$
\begin{bmatrix} \Delta x_t \\ \Delta Y_t \\ \Delta Z_t \end{bmatrix} = \begin{bmatrix} -0.2 & 0.1 & 0.1 \\ -0.16 & 0.08 & 0.08 \\ 0.36 & -0.18 & -0.18 \end{bmatrix} \begin{bmatrix} x_{t-1} \\ Y_{t-1} \\ Z_{t-1} \end{bmatrix} + \begin{bmatrix} \varepsilon_{1,t} \\ \varepsilon_{2,t} \\ \varepsilon_{3,t} \end{bmatrix}.
$$

The eigenvalues of π are 0, 0 and $-.3$, and so rank$(\pi) = 1$. Recall that the number of cointegrating vectors is the rank of π in a cointegrated system. In this example, there is one cointegrating vector, which can be computed by transforming π into row-echelon form,

$$
\begin{bmatrix} -0.2 & 0.1 & 0.1 \\ -0.16 & 0.08 & 0.08 \\ 0.36 & -0.18 & -0.18 \end{bmatrix} \Rightarrow \begin{bmatrix} 1 & -0.5 & -0.5 \\ -0.16 & 0.08 & 0.08 \\ 0.36 & -0.18 & -0.18 \end{bmatrix} \Rightarrow \begin{bmatrix} 1 & -0.5 & -0.5 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}
$$

The cointegrating vector is $\beta = [1 \ -0.5 \ -0.5]'$ and α is found by noting that

$$
\pi = \alpha \beta' = \begin{bmatrix} \alpha_1 & -\frac{1}{2}\alpha_1 & -\frac{1}{2}\alpha_1 \\ \alpha_2 & -\frac{1}{2}\alpha_2 & -\frac{1}{2}\alpha_2 \\ \alpha_3 & -\frac{1}{2}\alpha_3 & -\frac{1}{2}\alpha_3 \end{bmatrix},
$$

so that $\alpha = [-.2 \ -.16 \ 0.36]'$ is the first column of π.

5.A.4   Cointegrated with two cointegrating relationships

Consider the VAR(1),

$$
\begin{bmatrix} x_t \\ Y_t \\ Z_t \end{bmatrix} = \begin{bmatrix} 0.3 & 0.4 & 0.3 \\ 0.1 & 0.5 & 0.4 \\ 0.2 & 0.2 & 0.6 \end{bmatrix} \begin{bmatrix} x_{t-1} \\ Y_{t-1} \\ Z_{t-1} \end{bmatrix} + \begin{bmatrix} \varepsilon_{1,t} \\ \varepsilon_{2,t} \\ \varepsilon_{3,t} \end{bmatrix}.
$$

The eigenvalues of the parameter matrix are 1, .2+.1i and .2-.1i, which have complex moduli of 1, .223 and .223, respectively. The VECM form of this model is

$$
\begin{bmatrix} \Delta x_t \\ \Delta Y_t \\ \Delta Z_t \end{bmatrix} = \left(\begin{bmatrix} 0.3 & 0.4 & 0.3 \\ 0.1 & 0.5 & 0.4 \\ 0.2 & 0.2 & 0.6 \end{bmatrix} - \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \right) \begin{bmatrix} x_{t-1} \\ Y_{t-1} \\ Z_{t-1} \end{bmatrix} + \begin{bmatrix} \varepsilon_{1,t} \\ \varepsilon_{2,t} \\ \varepsilon_{3,t} \end{bmatrix}
$$

$$
\begin{bmatrix} \Delta x_t \\ \Delta Y_t \\ \Delta Z_t \end{bmatrix} = \begin{bmatrix} -0.7 & 0.4 & 0.3 \\ 0.1 & -0.5 & 0.4 \\ 0.2 & 0.2 & -0.4 \end{bmatrix} \begin{bmatrix} x_{t-1} \\ Y_{t-1} \\ Z_{t-1} \end{bmatrix} + \begin{bmatrix} \varepsilon_{1,t} \\ \varepsilon_{2,t} \\ \varepsilon_{3,t} \end{bmatrix}.
$$

The eigenvalues of π are 0, $-0.8+0.1i$ and $-0.8-0.1i$, and so $\text{rank}(\pi) = 2$. The number of cointegrating vectors is the rank of π. One set of cointegrating vectors can be found by transforming π into row-echelon form[22],

$$
\begin{bmatrix} -0.7 & 0.4 & 0.3 \\ 0.1 & -0.5 & 0.4 \\ 0.2 & 0.2 & -0.4 \end{bmatrix} \Rightarrow \begin{bmatrix} 1 & -0.57143 & -0.42857 \\ 0.1 & -0.5 & 0.4 \\ 0.2 & 0.2 & -0.4 \end{bmatrix} \Rightarrow \begin{bmatrix} 1 & -0.57143 & -0.42857 \\ 0 & -0.44286 & 0.44286 \\ 0 & 0.31429 & -0.31429 \end{bmatrix} \Rightarrow
$$

$$
\begin{bmatrix} 1 & -0.57143 & -0.42857 \\ 0 & 1 & -1 \\ 0 & 0.31429 & -0.31429 \end{bmatrix} \Rightarrow \begin{bmatrix} 1 & 0 & -1 \\ 0 & 1 & -1 \\ 0 & 0 & 0 \end{bmatrix}
$$

β is the transpose of first two rows of the row-echelon form,

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

α is found using the relationship

$$
\pi = \alpha\beta' = \begin{bmatrix} \alpha_{11} & \alpha_{12} & -\alpha_{11} - \alpha_{12} \\ \alpha_{21} & \alpha_{22} & -\alpha_{21} - \alpha_{22} \\ \alpha_{31} & \alpha_{32} & -\alpha_{31} - \alpha_{32} \end{bmatrix},
$$

and so α is the first two columns of π,

$$
\alpha = \begin{bmatrix} -0.7 & 0.4 \\ 0.1 & -0.5 \\ 0.2 & 0.2 \end{bmatrix}.
$$

---

[22]The cointegrating vectors are only defined up to an arbitrary normalization. Any set of cointegrating vectors β and be used to create a different set by multiplying by a k by k full-rank matrix $\mathbf{A}$ so that $\tilde{\beta} = \mathbf{A}\beta$ is also a cointegrating vector.

Exercises

Shorter Questions

Problem 5.1. Under what conditions are two random variables cointegrated?

Problem 5.2. Suppose $\mathbf{Y}_t = \Phi_0 + \Phi_1 \mathbf{Y}_{t-1} + \varepsilon_t$ where $\mathbf{Y}_t$ is a k by 1 vector values variable and Φ_0 and Φ_1 are conformable. What are the 1 and 2 step forecasts from this model?

Longer Questions

Exercise 5.1. Consider the estimation of a mean where the errors are a white noise process.

i. Show that the usual mean estimator is unbiased and derive its variance *without assuming the errors are i.i.d.*

   Now suppose error process follows an MA(1) so that $\varepsilon_t = v_t + \theta_1 v_{t-1}$ where v_t is a WN process.

ii. Show that the usual mean estimator is still unbiased and derive the variance of the mean.

   Suppose that $\{\eta_{1,t}\}$ and $\{\eta_{2,t}\}$ are two sequences of uncorrelated i.i.d. standard normal random variables.

$$x_t = \eta_{1,t} + \theta_{11}\eta_{1,t-1} + \theta_{12}\eta_{2,t-1}$$
$$Y_t = \eta_{2,t} + \theta_{21}\eta_{1,t-1} + \theta_{22}\eta_{2,t-1}$$

iii. What are $\mathrm{E}_t[x_{t+1}]$ and $\mathrm{E}_t[x_{t+2}]$?

iv. Define the autocovariance matrix of a vector process.

v. Compute the autocovariance matrix Γ_j for $j = 0, \pm 1$.

Exercise 5.2. Consider an AR(1)

i. What are the two types of stationarity? Provide precise definitions.

ii. Which of the following bivariate Vector Autoregressions are stationary? If they are not stationary are they cointegrated, independent unit roots or explosive? Assume

$$\begin{bmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{bmatrix} \overset{\text{i.i.d.}}{\sim} N(\mathbf{0}, \mathbf{I}_2)$$

Recall that the eigenvalues values of a 2×2 non-triangular matrix

$$\pi = \begin{bmatrix} \pi_{11} & \pi_{12} \\ \pi_{21} & \pi_{22} \end{bmatrix}$$

can be solved using the two-equation, two-unknowns system $\lambda_1 + \lambda_2 = \pi_{11} + \pi_{22}$ and $\lambda_1 \lambda_2 = \pi_{11}\pi_{22} - \pi_{12}\pi_{21}$.

(a)

$$\begin{bmatrix} x_t \\ Y_t \end{bmatrix} = \begin{bmatrix} 1.4 & .4 \\ -.6 & .4 \end{bmatrix} \begin{bmatrix} x_{t-1} \\ Y_{t-1} \end{bmatrix} + \begin{bmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{bmatrix}$$

(b)

$$\begin{bmatrix} x_t \\ Y_t \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} x_{t-1} \\ Y_{t-1} \end{bmatrix} + \begin{bmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{bmatrix}$$

(c)

$$\begin{bmatrix} x_t \\ Y_t \end{bmatrix} = \begin{bmatrix} .8 & 0 \\ .2 & .4 \end{bmatrix} \begin{bmatrix} x_{t-1} \\ Y_{t-1} \end{bmatrix} + \begin{bmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{bmatrix}$$

iii. What are spurious regression and balance?

iv. Why is spurious regression a problem?

v. Briefly outline the steps needed to test for a spurious regression in

$$Y_t = \beta_1 + \beta_2 x_t + \varepsilon_t.$$

Exercise 5.3. Consider the AR(2)

$$Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \varepsilon_t.$$

i. Rewrite the model with ΔY_t on the left-hand side and Y_{t-1} and ΔY_{t-1} on the right-hand side.

ii. What restrictions are needed on ϕ_1 and ϕ_2 for this model to collapse to an AR(1) in the first differences?

iii. When the model collapses, what does this imply about Y_t?

Consider the VAR(1)

$$x_t = x_{t-1} + \varepsilon_{1,t}$$
$$Y_t = \beta x_{t-1} + \varepsilon_{2,t}$$

where $\{\varepsilon_t\}$ is a vector white noise process.

i. Are x_t and Y_t cointegrated?

ii. Write this model in error correction form.

Consider the VAR(1)

$$\begin{bmatrix} x_t \\ Y_t \end{bmatrix} = \begin{bmatrix} 0.4 & 0.3 \\ 0.8 & 0.6 \end{bmatrix} \begin{bmatrix} x_{t-1} \\ Y_{t-1} \end{bmatrix} + \begin{bmatrix} \varepsilon_{1,t} \\ \varepsilon_{2,t} \end{bmatrix}$$

where $\{\varepsilon_t\}$ is a vector white noise process.

i. Verify that x_t and Y_t are cointegrated.

  ii. Write this model in error correction form.

  iii. Compute the speed of adjustment coefficient α and the cointegrating vector β where the β on x_t is normalized to 1.

Exercise 5.4. Data on interest rates on US government debt is collected for 3-month ($3MO$) T-bills, and 3-year ($3YR$) and 10-year ($10YR$) bonds from 1957 until 2009. Three transformed variables are defined using these three series:

Level	$3MO$
Slope	$10YR - 3MO$
Curvature	$(10YR - 3YR) - (3YR - 3MO)$

  i. In terms of VAR analysis, does it matter whether the original data or the level-slope-curvature model is fit? Hint: Think about reparameterizations between the two.

Granger Causality analysis is performed on this set, and the p-values are

	Level_{t-1}	Slope_{t-1}	Curvature_{t-1}
Level_t	0.000	0.244	0.000
Slope_t	0.000	0.000	0.000
Curvature_t	0.000	0.000	0.000
All (excl. self)	0.000	0.000	0.000

  ii. Interpret this table.

  iii. When constructing impulse response graphs the selection of the covariance of the shocks is important. Outline the alternatives and describe situations when each may be preferable.

  iv. Figure 5.7 contains the impulse response curves for this model. Interpret the graph. Also, comment on why the impulse responses can all be significantly different from 0 in light of the Granger Causality table.

  v. Why are some of the lag-0 impulses precisely 0.0?

Exercise 5.5. Answer the following questions:

  i. Consider the AR(2)
$$Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \varepsilon_t$$

    (a) Rewrite the model with ΔY_t on the left-hand side and Y_{t-1} and ΔY_{t-1} on the right-hand side.

    (b) What restrictions are needed on ϕ_1 and ϕ_2 for this model to collapse to an AR(1) in the first differences?

    (c) When the model collapses, what does this imply about Y_t?
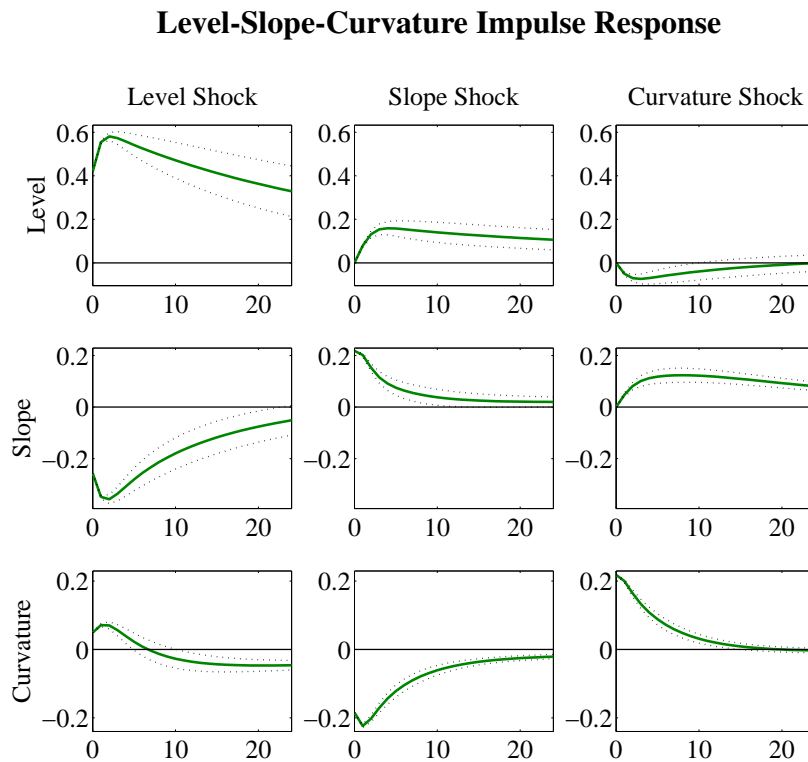
Level-Slope-Curvature Impulse Response

Figure 5.7: Impulse response functions and 95% confidence intervals for the level-slope-curvature exercise.

---

ii. Consider the VAR(1)

$$x_t = x_{t-1} + \varepsilon_{1,t}$$
$$Y_t = \beta x_{t-1} + \varepsilon_{2,t}$$

where $\{\varepsilon_t\}$ is a vector white noise process.

(a) Are x_t and Y_t cointegrated?

(b) Write this model in error correction form.

iii. Consider the VAR(1)

$$\begin{bmatrix} x_t \\ Y_t \end{bmatrix} = \begin{bmatrix} 0.625 & -0.3125 \\ -0.75 & 0.375 \end{bmatrix} \begin{bmatrix} x_{t-1} \\ Y_{t-1} \end{bmatrix} + \begin{bmatrix} \varepsilon_{1,t} \\ \varepsilon_{2,t} \end{bmatrix}$$

where $\{\varepsilon_t\}$ is a vector white noise process.

(a) Verify that x_t and Y_t are cointegrated.

(b) Write this model in error correction form.

(c) Compute the speed of adjustment coefficient α and the cointegrating vector β where the β on x_t is normalized to 1.

Exercise 5.6. Consider the estimation of a mean where the errors are a white noise process.

i. Show that the usual mean estimator is unbiased and derive its variance *without assuming the errors are i.i.d.*

Now suppose error process follows an AR(1) so that $Y_t = \mu + \varepsilon_t$ and $\varepsilon_t = \rho \varepsilon_{t-1} + V_t$ where $\{V_t\}$ is a WN process.

ii. Show that the usual mean estimator is still unbiased and derive the variance of the sample mean.

iii. What is Granger Causality and how is it useful in Vector Autoregression analysis? Be as specific as possible.

Suppose that $\{\eta_{1,t}\}$ and $\{\eta_{2,t}\}$ are two sequences of uncorrelated i.i.d. standard normal random variables.

$$x_t = \eta_{1,t} + \theta_{11}\eta_{1,t-1} + \theta_{12}\eta_{2,t-1}$$
$$Y_t = \eta_{2,t} + \theta_{21}\eta_{1,t-1} + \theta_{22}\eta_{2,t-1}$$

iv. Define the autocovariance matrix of a vector process.

v. Compute the autocovariance matrix Γ_j for $j = 0, \pm 1$.

vi. The AIC, HQIC, and BIC are computed for a bivariate VAR with lag length ranging from 0 to 12 and are in the table below. Which model is selected by each criterion?

Lag Length	AIC	HQIC	BIC
0	2.1916	2.1968	2.2057
1	0.9495	0.9805	1.0339
2	0.9486	1.0054	1.1032
3	0.9716	1.0542	1.1965
4	0.9950	1.1033	1.2900
5	1.0192	1.1532	1.3843
6	1.0417	1.2015	1.4768
7	1.0671	1.2526	1.5722
8	1.0898	1.3010	1.6649
9	1.1115	1.3483	1.7564
10	1.1331	1.3956	1.8478
11	1.1562	1.4442	1.9406
12	1.1790	1.4926	2.0331

Exercise 5.7. Consider the VAR(1)

$$x_t = x_{t-1} + \varepsilon_{1,t}$$
$$Y_t = \beta x_{t-1} + \varepsilon_{2,t}$$

where $\{\varepsilon_t\}$ is a vector white noise process.

  i. Are x_t and Y_t cointegrated?

 ii. Write this model in error correction form.

Exercise 5.8. Answer the following questions.

  i. Describe two methods for determining the number of lags to use in a VAR(P)

 ii. Consider the VAR(P)
$$\mathbf{Y_t} = \Phi_1 \mathbf{Y_{t-1}} + \Phi_2 \mathbf{Y_{t-2}} + \varepsilon_t.$$

 Write this VAR in companion form. Under what conditions is this process stationary?

iii. For the remainder of the question, consider the 2-dimentional VAR(1)

$$\mathbf{Y_t} = \Phi_1 \mathbf{Y_{t-1}} + \varepsilon_t.$$

 Define Granger Causality and explain what conditions on Φ_1 are needed for no series in $\mathbf{y}_t$ to Granger cause any other series in $\mathbf{y}_t$.

 iv. Define cointegration in this system.

  v. What conditions on Φ_1 are required for the VAR(1) to be cointegrated?

 vi. Write the VAR(1) in error correction form.

vii. In this setup, describe how to test for cointegration using the Engle-Granger method.

Exercise 5.9. Consider a VAR(1)
$$\mathbf{Y}_t = \Phi_1 \mathbf{Y}_{t-1} + \varepsilon_t$$

  i. What are the impulses in this model?

 ii. Define cointegration for this model.

iii. What conditions on the eigenvalues of Φ_1 are required for cointegration to be present?

 iv. Consider a 2-dimensional VAR(1) written in error correction form

$$\Delta \mathbf{Y}_t = \Pi \mathbf{Y}_{t-1} + \varepsilon_t.$$

 Assume each of the variables in $\mathbf{Y}_t$ are $I(1)$. What conditions on the rank of Π must hold when:

 (a) $\mathbf{Y}_{t-1}$ are stationary
 (b) $\mathbf{Y}_{t-1}$ are cointegrated

(c) $\mathbf{Y}_{t-1}$ are random walks

v. Define spurious regression. Why is this a problem?

Exercise 5.10. Consider the VAR(P)

$$\mathbf{Y}_t = \Phi_1 \mathbf{Y}_{t-1} + \Phi_2 \mathbf{Y}_{t-2} + \varepsilon_t$$

i. Write this in companion form. Under what conditions is the VAR(P) stationary?

ii. Consider the 2-dimentional VAR

$$\mathbf{Y}_t = \Phi_1 \mathbf{Y}_{t-1} + \varepsilon_t$$

(a) What conditions on Φ_1 are required for the VAR(1) to have cointegration?

(b) Describe how to test for cointegration using the Engle-Granger method.

Exercise 5.11. Consider a VAR(1)

$$\mathbf{Y}_t = \Phi_1 \mathbf{Y}_{t-1} + \varepsilon_t$$

i. What is an impulse response function for this model?

ii. Define cointegration for this model.

iii. What conditions on the eigenvalues of Φ_1 are required for cointegration to be present?

iv. Consider a 2-dimensional VAR(1) written in error correction form

$$\Delta \mathbf{Y}_t = \Pi \mathbf{Y}_{t-1} + \varepsilon_t.$$

Assume each of the variables in $\mathbf{Y}_t$ are $I(1)$. What conditions on the rank of Π must hold when:

(a) $\mathbf{Y}_{t-1}$ are stationary

(b) $\mathbf{Y}_{t-1}$ are cointegrated

(c) $\mathbf{Y}_{t-1}$ are random walks

v. Define spurious regression. Why is this a problem?

Exercise 5.12. Answer the following questions.

i. Describe two methods for determining the number of lags to use in a VAR(P)

ii. Consider the VAR(P)

$$\mathbf{Y}_t = \Phi_1 \mathbf{Y}_{t-1} + \Phi_2 \mathbf{Y}_{t-2} + \varepsilon_t.$$

Write this in companion form. Under what conditions is the VAR(P) stationary?

iii. For the remainder of the question, consider the 2-dimentional VAR(1)

$$\mathbf{Y}_t = \Phi_1 \mathbf{Y}_{t-1} + \varepsilon_t.$$

Define Granger Causality and explain what conditions on Φ_1 are needed for no series in $\mathbf{Y}_t$ to Granger cause any other series in $\mathbf{Y}_t$.

iv. Define cointegration in this system.

v. What conditions on Φ_1 are required for the VAR(1) to have cointegration?

vi. Write the VAR(1) in error correction form.

vii. In this setup, describe how to test for cointegration using the Engle-Granger method.

Exercise 5.13. Answer the following questions.

i. Suppose $Y_t = \phi_0 + \phi_1 y + \phi_2 Y_{t-2} + \varepsilon_t$ where $\{\varepsilon_t\}$ is a white noise process.

ii. Write this model in companion form.

   (a) Using the companion form, derive expressions for the first two autocovariances of Y_t. (It is not necessary to explicitly solve them in scalar form)

   (b) Using the companion form, determine the formal conditions for ϕ_1 and ϕ_2 to for $\{Y_t\}$ to be covariance stationary. You can use the result that when $\mathbf{A}$ is a 2 by 2 matrix, its eigenvalues solve the two equations

$$\begin{aligned}
\lambda_1 \lambda_2 &= a_{11}a_{22} - a_{12}a_{21} \\
\lambda_1 + \lambda_2 &= a_{11} + a_{22}
\end{aligned}$$