Chapter 5

Analysis of Multiple Time Series

Note: The primary references for these notes are chapters 5 and 6 in Enders (2004). An alternative, but more technical treatment can be found in chapters 10-11 and 18-19 in Hamilton (1994).

Multivariate time-series analysis extends many of the ideas of univariate time-series analysis to systems of equations. The primary model in multivariate time-series analysis is the vector autoregression (VAR), a direct and natural extension of the univariate autoregression. Most results that apply to univariate time-series can be directly ported to multivariate time-series with a slight change in notation and the use of linear algebra. The chapter examines both stationary and nonstationary vector processes through VAR analysis, cointegration and spurious regression. This chapter discusses properties of vector time-series models, estimation and identification as well as Granger causality and Impulse Response Functions. The chapter concludes by examining the contemporaneous relationship between two or more time-series in the framework of cointegration, spurious regression and cross-sectional regression of stationary time-series.

In many situations, analyzing a time-series in isolation is reasonable; in other cases univariate analysis may be limiting. For example, Campbell (1996) links financially interesting variables, including stock returns and the default premium, in a multivariate system that allows shocks to one variable to propagate to the others. The vector autoregression is the mechanism that is used to link multiple *stationary* time-series variables together. When variables contain unit roots, a different type of analysis, cointegration, is needed. This chapter covers these two topics building on many results from the analysis of univariate time-series.

5.1 Vector Autoregressions

Vector autoregressions are remarkably similar to univariate autoregressions; so similar that the intuition behind most results carries over by simply replacing scalars with matrices and scalar operations with matrix operations.

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 P[th] order vector autoregression , written VAR(P), is a process that evolves according to

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

where $\mathbf{y}_t$ is a k by 1 vector stochastic process, $\mathbf{\Phi}_0$ is a k by 1 vector of intercept parameters, $\mathbf{\Phi}_j$, $j = 1, \ldots, P$ are k by k parameter matrices and $\boldsymbol{\epsilon}_t$ is a vector white noise process with the additional assumption that $\mathrm{E}_{t-1}[\boldsymbol{\epsilon}_t] = \mathbf{0}$.

Simply replacing the vectors and matrices with scalars will produce the definition of an AR(P). A vector white noise process has the same useful properties as a univariate white noise process; it is mean zero, has finite covariance and is uncorrelated with its past although the elements of a vector white noise process are not required to be *contemporaneously* uncorrelated.

Definition 5.2 (Vector White Noise Process)**.** A k by 1 vector valued stochastic process, $\{\boldsymbol{\epsilon}_t\}$is said to be a vector white noise if

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

where $\mathbf{\Sigma}$ is a finite positive definite matrix.

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

$$\mathbf{y}_t = \mathbf{\Phi}_0 + \mathbf{\Phi}_1 \mathbf{y}_{t-1} + \boldsymbol{\epsilon}_t,$$

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

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

It is clear that each element of $\mathbf{y}_t$ is a function of each element of $\mathbf{y}_{t-1}$, although certain parameterizations of $\mathbf{\Phi}_1$ may remove the dependence. Treated as individual time-series, deriving the properties of VARs is an exercise in tedium. However, a few tools from linear algebra make working with VARs hardly more difficult than autoregressions.

5.1.2 Properties of a VAR(1)

The properties of the VAR(1) are fairly simple to study. More importantly, section 5.2 shows that all VAR(P)s can be rewritten as a VAR(1), and so the general case requires no additional effort than the first order VAR.

5.1.2.1 Stationarity

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

$$\mathbf{y}_t = \mathbf{\Phi}_0 + \mathbf{\Phi}_1 \mathbf{y}_{t-1} + \boldsymbol{\epsilon}_t$$

is covariance stationary if the eigenvalues of $\mathbf{\Phi}_1$ are less than 1 in modulus.[1] In the univariate case, this is equivalent to the condition $|\phi_1| < 1$. Assuming the eigenvalues of $\boldsymbol{\phi}_1$ are less than one in absolute value, backward substitution can be used to show that

$$\mathbf{y}_t = \sum_{i=0}^{\infty} \mathbf{\Phi}_1^i \mathbf{\Phi}_0 + \sum_{i=0}^{\infty} \mathbf{\Phi}_1^i \boldsymbol{\epsilon}_{t-i} \tag{5.3}$$

which, applying Theorem 5.1, is equivalent to

$$\mathbf{y}_t = (\mathbf{I}_k - \mathbf{\Phi}_1)^{-1} \mathbf{\Phi}_0 + \sum_{i=0}^{\infty} \mathbf{\Phi}_1^i \boldsymbol{\epsilon}_{t-i} \tag{5.4}$$

where the eigenvalue condition ensures that $\mathbf{\Phi}_1^i$ will converge to zero as i grows large.

5.1.2.2 Mean

Taking expectations of $\mathbf{y}_t$ using the backward substitution form yields

[1] The definition of an eigenvalue is:

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.

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

Theorem 5.1 (Matrix Power)**.** *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).

$$\begin{aligned}
\mathrm{E}\left[\mathbf{y}_t\right] &= \mathrm{E}\left[\left(\mathbf{I}_k - \boldsymbol{\Phi}_1\right)^{-1}\boldsymbol{\Phi}_0\right] + \mathrm{E}\left[\sum_{i=0}^{\infty}\boldsymbol{\Phi}_1^i\boldsymbol{\epsilon}_{t-i}\right] \\
&= \left(\mathbf{I}_k - \boldsymbol{\Phi}_1\right)^{-1}\boldsymbol{\Phi}_0 + \sum_{i=0}^{\infty}\boldsymbol{\Phi}_1^i\mathrm{E}\left[\boldsymbol{\epsilon}_{t-i}\right] \\
&= \left(\mathbf{I}_k - \boldsymbol{\Phi}_1\right)^{-1}\boldsymbol{\Phi}_0 + \sum_{i=0}^{\infty}\boldsymbol{\Phi}_1^i\mathbf{0} \\
&= \left(\mathbf{I}_k - \boldsymbol{\Phi}_1\right)^{-1}\boldsymbol{\Phi}_0
\end{aligned} \tag{5.5}$$

This result is similar to that of a univariate AR(1) which has a mean of $(1-\phi_1)^{-1}\phi_0$. The eigenvalues play an important role in determining the mean. If an eigenvalue of $\boldsymbol{\Phi}_1$ is close to one, $(\mathbf{I}_k - \boldsymbol{\Phi}_1)^{-1}$ will contain large values and the unconditional mean will be large. Similarly, if $\boldsymbol{\Phi}_1 = \mathbf{0}$, then the mean is $\boldsymbol{\Phi}_0$ since $\{y_t\}$ is composed of white noise and a constant.

5.1.2.3 Variance

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

$$\mathbf{y}_t = \boldsymbol{\Phi}_0 + \boldsymbol{\Phi}_1\mathbf{y}_{t-1} + \boldsymbol{\Phi}_2\mathbf{y}_{t-2} + \ldots + \boldsymbol{\Phi}_P\mathbf{y}_{t-P} + \boldsymbol{\epsilon}_t$$

is given by

$$\begin{aligned}
\mathbf{y}_t - \boldsymbol{\mu} &= \boldsymbol{\Phi}_1\left(\mathbf{y}_{t-1} - \boldsymbol{\mu}\right) + \boldsymbol{\Phi}_2\left(\mathbf{y}_{t-2} - \boldsymbol{\mu}\right) + \ldots + \boldsymbol{\Phi}_P\left(\mathbf{y}_{t-P} - \boldsymbol{\mu}\right) + \boldsymbol{\epsilon}_t \\
\tilde{\mathbf{y}}_t &= \boldsymbol{\Phi}_1\tilde{\mathbf{y}}_{t-1} + \boldsymbol{\Phi}_2\tilde{\mathbf{y}}_{t-2} + \ldots + \boldsymbol{\Phi}_P\tilde{\mathbf{y}}_{t-P} + \boldsymbol{\epsilon}_t
\end{aligned} \tag{5.6}$$

and in the case of a VAR(1),

$$\tilde{\mathbf{y}}_t = \sum_{i=1}^{\infty}\boldsymbol{\phi}_1^i\boldsymbol{\epsilon}_{t-i} \tag{5.7}$$

The deviations form is simply a translation of the VAR from its original mean, $\boldsymbol{\mu}$, to a mean of $\mathbf{0}$. The advantage of the deviations form is that all dynamics *and* the shocks are identical, and so can be used in deriving the long-run covariance, autocovariances and in forecasting. Using the backward substitution form of a VAR(1), the long run covariance can be derived as

$$
\begin{aligned}
\mathrm{E}\left[(\mathbf{y}_t - \boldsymbol{\mu})(\mathbf{y}_t - \boldsymbol{\mu})'\right] = \mathrm{E}\left[\tilde{\mathbf{y}}_t \tilde{\mathbf{y}}_t'\right] &= \mathrm{E}\left[\left(\sum_{i=0}^{\infty} \boldsymbol{\Phi}_1^i \boldsymbol{\epsilon}_{t-i}\right)\left(\sum_{i=0}^{\infty} \boldsymbol{\Phi}_1^i \boldsymbol{\epsilon}_{t-i}\right)'\right] \\
&= \mathrm{E}\left[\sum_{i=0}^{\infty} \boldsymbol{\Phi}_1^i \boldsymbol{\epsilon}_{t-i} \boldsymbol{\epsilon}_{t-i}' \left(\boldsymbol{\Phi}_1'\right)^i\right] \quad \text{(Since } \boldsymbol{\epsilon}_t \text{ is WN)} \\
&= \sum_{i=0}^{\infty} \boldsymbol{\Phi}_1^i \mathrm{E}\left[\boldsymbol{\epsilon}_{t-i} \boldsymbol{\epsilon}_{t-i}'\right] \left(\boldsymbol{\Phi}_1'\right)^i \\
&= \sum_{i=0}^{\infty} \boldsymbol{\Phi}_1^i \boldsymbol{\Sigma} \left(\boldsymbol{\Phi}_1'\right)^i \\
\operatorname{vec}\left(\mathrm{E}\left[(\mathbf{y}_t - \boldsymbol{\mu})(\mathbf{y}_t - \boldsymbol{\mu})'\right]\right) &= \left(\mathbf{I}_{k^2} - \boldsymbol{\Phi}_1 \otimes \boldsymbol{\Phi}_1\right)^{-1} \operatorname{vec}(\boldsymbol{\Sigma})
\end{aligned}
\tag{5.8}
$$

where $\boldsymbol{\mu} = (\mathbf{I}_k - \boldsymbol{\Phi}_1)^{-1}\boldsymbol{\Phi}_0$. Compared to the long-run variance of a univariate autoregression, $\sigma^2/(1-\phi_1^2)$, the similarities are less obvious. The differences arise from the noncommutative nature of matrices ($\mathbf{AB} \neq \mathbf{BA}$ in general). The final line makes use of the *vec* (vector) operator to re-express the covariance. The vec operator and a Kronecker product stack the elements of a matrix product into a single column.[2]

Once again the eigenvalues of $\boldsymbol{\Phi}_1$ play an important role. If any are close to 1, the variance will be large since the eigenvalues fundamentally determine the persistence of shocks: as was

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

Definition 5.4 (vec)**.** Let $\mathbf{A} = [a_{ij}]$ be an m by n matrix. The vec operator (also known as the *stack* operator) is defined

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

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

The Kronecker Product is defined:

Definition 5.5 (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 5.2 (Kronecker and vec of a product)**.** *Let* $\mathbf{A}$, $\mathbf{B}$ *and* $\mathbf{C}$ *be conformable matrices as needed. Then*

$$
\operatorname{vec}(\mathbf{ABC}) = \left(\mathbf{C}' \otimes \mathbf{A}\right) \operatorname{vec} \mathbf{B}
$$

the case in scalar autoregressions, higher persistence lead to larger variances.

5.1.2.4 Autocovariance

The autocovariances of a vector valued stochastic process are defined

Definition 5.6 (Autocovariance)**.** The autocovariance matrices of k by 1 valued vector covariance stationary stochastic process $\{\mathbf{y}_t\}$ are defined

$$\mathbf{\Gamma}_s = \mathrm{E}[(\mathbf{y}_t - \boldsymbol{\mu})(\mathbf{y}_{t-s} - \boldsymbol{\mu})'] \tag{5.10}$$

and

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

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

These present the first significant deviation from the univariate time-series analysis in chapter 4. Instead of being symmetric around t, they are symmetric in their transpose. Specifically,

$$\mathbf{\Gamma}_s \neq \mathbf{\Gamma}_{-s}$$

but it is the case that[3]

$$\mathbf{\Gamma}_s = \mathbf{\Gamma}_{-s}'.$$

In contrast, the autocovariances of stationary scalar processes satisfy $\gamma_s = \gamma_{-s}$. Computing the autocovariances is also easily accomplished using the backward substitution form,

$$\begin{aligned}
\mathbf{\Gamma}_s = \mathrm{E}\left[(\mathbf{y}_t - \boldsymbol{\mu})(\mathbf{y}_{t-s} - \boldsymbol{\mu})'\right] &= \mathrm{E}\left[\left(\sum_{i=0}^{\infty} \mathbf{\Phi}_1^i \boldsymbol{\epsilon}_{t-i}\right)\left(\sum_{i=0}^{\infty} \mathbf{\Phi}_1^i \boldsymbol{\epsilon}_{t-s-i}\right)'\right] && (5.12)\\
&= \mathrm{E}\left[\left(\sum_{i=0}^{s-1} \mathbf{\Phi}_1^i \boldsymbol{\epsilon}_{t-i}\right)\left(\sum_{i=0}^{\infty} \mathbf{\Phi}_1^i \boldsymbol{\epsilon}_{t-s-i}\right)'\right] \\
&\quad + \mathrm{E}\left[\left(\sum_{i=0}^{\infty} \mathbf{\Phi}_1^s \mathbf{\Phi}_1^i \boldsymbol{\epsilon}_{t-s-i}\right)\left(\sum_{i=0}^{\infty} \mathbf{\Phi}_1^i \boldsymbol{\epsilon}_{t-s-i}\right)'\right] && (5.13)\\
&= \mathbf{0} + \mathbf{\Phi}_1^s \mathrm{E}\left[\left(\sum_{i=0}^{\infty} \mathbf{\Phi}_1^i \boldsymbol{\epsilon}_{t-s-i}\right)\left(\sum_{i=0}^{\infty} \mathbf{\Phi}_1^i \boldsymbol{\epsilon}_{t-s-i}\right)'\right] \\
&= \mathbf{\Phi}_1^s \mathrm{V}[\mathbf{y}_t]
\end{aligned}$$

and

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

$$\begin{aligned}
\boldsymbol{\Gamma}_{-s} = \mathrm{E}\left[(\mathbf{y}_t - \boldsymbol{\mu})(\mathbf{y}_{t+s} - \boldsymbol{\mu})'\right] &= \mathrm{E}\left[\left(\sum_{i=0}^{\infty} \boldsymbol{\Phi}_1^i \boldsymbol{\epsilon}_{t-i}\right)\left(\sum_{i=0}^{\infty} \boldsymbol{\Phi}_1^i \boldsymbol{\epsilon}_{t+s-i}\right)'\right] \qquad (5.14)\\
&= \mathrm{E}\left[\left(\sum_{i=0}^{\infty} \boldsymbol{\Phi}_1^i \boldsymbol{\epsilon}_{t-i}\right)\left(\sum_{i=0}^{\infty} \boldsymbol{\Phi}_1^s \boldsymbol{\Phi}_1^i \boldsymbol{\epsilon}_{t-i}\right)'\right]\\
&\quad + \mathrm{E}\left[\left(\sum_{i=0}^{\infty} \boldsymbol{\Phi}_1^i \boldsymbol{\epsilon}_{t-i}\right)\left(\sum_{i=0}^{s-1} \boldsymbol{\Phi}_1^i \boldsymbol{\epsilon}_{t+s-i}\right)'\right] \qquad (5.15)\\
&= \mathrm{E}\left[\left(\sum_{i=0}^{\infty} \boldsymbol{\Phi}_1^i \boldsymbol{\epsilon}_{t-i}\right)\left(\sum_{i=0}^{\infty} \boldsymbol{\epsilon}_{t-i}' \left(\boldsymbol{\Phi}_1'\right)^i \left(\boldsymbol{\Phi}_1'\right)^s\right)\right] + \mathbf{0}\\
&= \mathrm{E}\left[\left(\sum_{i=0}^{\infty} \boldsymbol{\Phi}_1^i \boldsymbol{\epsilon}_{t-i}\right)\left(\sum_{i=0}^{\infty} \boldsymbol{\epsilon}_{t-i}' \left(\boldsymbol{\Phi}_1'\right)^i\right)\right]\left(\boldsymbol{\Phi}_1'\right)^s\\
&= \mathrm{V}\left[\mathbf{y}_t\right]\left(\boldsymbol{\Phi}_1'\right)^s
\end{aligned}$$

where $\mathrm{V}[\mathbf{y}_t]$ is the symmetric covariance matrix of the VAR. Like most properties of a VAR, this result is similar to the autocovariance function of an AR(1): $\gamma_s = \phi_1^s \sigma^2/(1-\phi_1^2) = \phi_1^s \mathrm{V}[y_t]$.

5.2 Companion Form

Once the properties of a VAR(1) have been studied, one surprising and useful result is that any stationary VAR(P) can be rewritten as a VAR(1). Suppose $\{\mathbf{y}_t\}$ follows a VAR(P) process,

$$\mathbf{y}_t = \boldsymbol{\Phi}_0 + \boldsymbol{\Phi}_1 \mathbf{y}_{t-1} + \boldsymbol{\Phi}_2 \mathbf{y}_{t-2} + \ldots + \boldsymbol{\Phi}_P \mathbf{y}_{t-P} + \boldsymbol{\epsilon}_t.$$

By subtracting the mean and stacking P of $\mathbf{y}_t$ into a large column vector denoted $\mathbf{z}_t$, a VAR(P) can be transformed into a VAR(1) by constructing the companion form.

Definition 5.7 (Companion Form of a VAR(P))**.** Let $\mathbf{y}_t$ follow a VAR(P) given by

$$\mathbf{y}_t = \boldsymbol{\Phi}_0 + \boldsymbol{\Phi}_1 \mathbf{y}_{t-1} + \boldsymbol{\Phi}_2 \mathbf{y}_{t-2} + \ldots + \boldsymbol{\Phi}_P \mathbf{y}_{t-P} + \boldsymbol{\epsilon}_t$$

where $\boldsymbol{\epsilon}_t$ is a vector white noise process and $\boldsymbol{\mu} = \left(\mathbf{I} - \sum_{p=1}^{P} \boldsymbol{\Phi}_p\right)^{-1} \boldsymbol{\Phi}_0 = \mathrm{E}[\mathbf{y}_t]$ is finite. The companion form is given by

$$\mathbf{z}_t = \boldsymbol{\Upsilon} \mathbf{z}_{t-1} + \boldsymbol{\xi}_t \qquad (5.16)$$

where

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

$$\boldsymbol{\Upsilon} = \begin{bmatrix} \boldsymbol{\Phi}_1 & \boldsymbol{\Phi}_2 & \boldsymbol{\Phi}_3 & \dots & \boldsymbol{\Phi}_{P-1} & \boldsymbol{\Phi}_P \\ \mathbf{I}_k & \mathbf{0} & \mathbf{0} & \dots & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_k & \mathbf{0} & \dots & \mathbf{0} & \mathbf{0} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \dots & \mathbf{I}_k & \mathbf{0} \end{bmatrix} \tag{5.18}$$

and

$$\xi_t = \begin{bmatrix} \boldsymbol{\epsilon}_t \\ \mathbf{0} \\ \vdots \\ \mathbf{0} \end{bmatrix}. \tag{5.19}$$

This is known as the **companion form** and allows the statistical properties of any VAR(P) to be directly computed using only the results of a VAR(1) noting that

$$\mathrm{E}[\xi_t \xi_t'] = \begin{bmatrix} \boldsymbol{\Sigma} & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \dots & \mathbf{0} \\ \vdots & \vdots & \vdots & \vdots \\ \mathbf{0} & \mathbf{0} & \dots & \mathbf{0} \end{bmatrix}.$$

Using this form, it can be determined that a VAR(P) is covariance stationary if all of the eigenvalues of $\boldsymbol{\Upsilon}$ - there are kP of them - are less than one in absolute value (modulus if complex).[4]

5.3 Empirical Examples

Throughout this chapter two examples from the finance literature will be used.

5.3.1 Example: The interaction of stock and bond returns

Stocks and long term 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)

$$\begin{bmatrix} VWM_t \\ 10YR_t \end{bmatrix} = \begin{bmatrix} \phi_{01} \\ \phi_{02} \end{bmatrix} + \begin{bmatrix} \phi_{11,1} & \phi_{12,1} \\ \phi_{21,1} & \phi_{22,1} \end{bmatrix} \begin{bmatrix} VWM_{t-1} \\ 10YR_{t-1} \end{bmatrix} + \begin{bmatrix} \epsilon_{1,t} \\ \epsilon_{2,t} \end{bmatrix}$$

[4] Companion form is also useful when working with univariate AR(P) models. An AR(P) can be reexpressed using its companion VAR(1) which allows properties such as the long-run variance and autocovariances to be easily computed.

which implies a model for stock returns

$$VWM_t = \phi_{01} + \phi_{11,1} VWM_{t-1} + \phi_{12,1} 10YR_{t-1} + \epsilon_{1,t}$$

and a model for long bond returns

$$10YR_t = \phi_{01} + \phi_{21,1} VWM_{t-1} + \phi_{22,1} 10YR_{t-1} + \epsilon_{2,t}.$$

Since these models do not share any parameters, they can be estimated separately using OLS. Using annualized return data for the VWM from CRSP and the 10-year constant maturity treasury yield from FRED covering the period May 1953 until December 2008, a VAR(1) was estimated.[5]

$$\begin{bmatrix} VWM_t \\ 10YR_t \end{bmatrix} = \begin{bmatrix} 9.733 \\ (0.000) \\ 1.058 \\ (0.000) \end{bmatrix} + \begin{bmatrix} 0.097 & 0.301 \\ (0.104) & (0.000) \\ -0.095 & 0.299 \\ (0.000) & (0.000) \end{bmatrix} \begin{bmatrix} VWM_{t-1} \\ 10YR_{t-1} \end{bmatrix} + \begin{bmatrix} \epsilon_{1,t} \\ \epsilon_{2,t} \end{bmatrix}$$

where the p-val is in parenthesis below each coefficient. A few things are worth noting. Stock returns are not predictable with their own lag but do appear to be predictable using lagged bond returns: positive bond returns lead to positive future returns in stocks. In contrast, positive returns in equities result in negative returns for future bond holdings. The long-run mean can be computed as

$$\left(\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} - \begin{bmatrix} 0.097 & 0.301 \\ -0.095 & 0.299 \end{bmatrix} \right)^{-1} \begin{bmatrix} 9.733 \\ 1.058 \end{bmatrix} = \begin{bmatrix} 10.795 \\ 0.046 \end{bmatrix}.$$

These values are similar to the sample means of 10.801 and 0.056.

5.3.2 Example: Campbell's VAR

Campbell (1996) builds a theoretical model for asset prices where economically meaningful variables evolve according to a VAR. Campbell's model included stock returns, real labor income growth, the term premium, the relative t-bill rate and the dividend yield. The VWM series from CRSP is used for equity returns. Real labor income is computed as the log change in income from labor minus the log change in core inflation and both series are from FRED. The term premium is the difference between the yield on a 10-year constant maturity bond and the 3-month t-bill rate. Both series are from FRED. The relative t-bill rate is the current yield on a 1-month t-bill minus the average yield over the past 12 months and the data is available on Ken French's web site. The dividend yield was computed as the difference in returns on the VWM with and without dividends; both series are available from CRSP.

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

[5] The yield is first converted to prices and then returns are computed as the log difference in consecutive prices.

Raw Data

	VWM_{t-1}	LBR_{t-1}	RTB_{t-1}	$TERM_{t-1}$	DIV_{t-1}
VWM_t	0.073 (0.155)	0.668 (0.001)	−0.050 (0.061)	−0.000 (0.844)	0.183 (0.845)
LBR_t	0.002 (0.717)	−0.164 (0.115)	0.002 (0.606)	0.000 (0.139)	−0.060 (0.701)
RTB_t	0.130 (0.106)	0.010 (0.974)	0.703 (0.000)	−0.010 (0.002)	0.137 (0.938)
$TERM_t$	−0.824 (0.084)	−2.888 (0.143)	0.069 (0.803)	0.960 (0.000)	4.028 (0.660)
DIV_t	0.001 (0.612)	−0.000 (0.989)	−0.001 (0.392)	−0.000 (0.380)	−0.045 (0.108)

Standardized Series

	VWM_{t-1}	LBR_{t-1}	RTB_{t-1}	$TERM_{t-1}$	DIV_{t-1}
VWM_t	0.073 (0.155)	0.112 (0.001)	−0.113 (0.061)	−0.011 (0.844)	0.007 (0.845)
LBR_t	0.012 (0.717)	−0.164 (0.115)	0.027 (0.606)	0.065 (0.139)	−0.013 (0.701)
RTB_t	0.057 (0.106)	0.001 (0.974)	0.703 (0.000)	−0.119 (0.002)	0.002 (0.938)
$TERM_t$	−0.029 (0.084)	−0.017 (0.143)	0.006 (0.803)	0.960 (0.000)	0.005 (0.660)
DIV_t	0.024 (0.612)	−0.000 (0.989)	−0.043 (0.392)	−0.043 (0.380)	−0.045 (0.108)

Table 5.1: Parameter estimates from Campbell's VAR. The top panel contains estimates using unscaled data while the bottom panel contains estimates from data which have been standardized to have unit variance. While the magnitudes of many coefficients change, the p-vals and the eigenvalues of these two parameter matrices are identical, and the parameters are roughly comparable since the series have the same variance.

$$\begin{bmatrix} VWM_t \\ LBR_t \\ RTB_t \\ TERM_t \\ DIV_t \end{bmatrix} = \mathbf{\Phi}_0 + \mathbf{\Phi}_1 \begin{bmatrix} VWM_{t-1} \\ LBR_{t-1} \\ RTB_{t-1} \\ TERM_{t-1} \\ DIV_{t-1} \end{bmatrix} + \begin{bmatrix} \epsilon_{1,t} \\ \epsilon_{2,t} \\ \epsilon_{3,t} \\ \epsilon_{4,t} \\ \epsilon_{5t} \end{bmatrix}.$$

Two sets of parameters are presented in table 5.1. The top panel contains estimates using non-scaled data. This produces some very large (in magnitude, not statistical significance) estimates which are the result of two variables having very different scales. The bottom panel contains estimates from data which have been standardized by dividing each series by its standard deviation. This makes the magnitude of all coefficients approximately comparable. Despite this transformation and very different parameter estimates, the p-vals remain unchanged. This shouldn't be surprising since OLS t-stats are invariant to scalings of this type. One less obvious

feature of the two sets of estimates is that the eigenvalues of the two parameter matrices are identical and so both sets of parameter estimates indicate the same persistence.

5.4 VAR forecasting

Once again, the behavior of a VAR(P) is identical to that of an AR(P). Recall that the h-step ahead forecast, $\hat{y}_{t+h|t}$ from an AR(1) is given by

$$\mathrm{E}_t[y_{t+h}] = \sum_{j=0}^{h-1} \phi_1^j \phi_0 + \phi_1^h y_t.$$

The h-step ahead forecast of a VAR(1) , $\hat{\mathbf{y}}_{t+h|t}$ is

$$\mathrm{E}_t[\mathbf{y}_{t+h}] = \sum_{j=0}^{h-1} \mathbf{\Phi}_1^j \mathbf{\Phi}_0 + \mathbf{\Phi}_1^h \mathbf{y}_t$$

Forecasts from higher order VARs can be constructed by direct forward recursion beginning at $h = 1$, although they are often simpler to compute using the deviations form of the VAR since it includes no intercept,

$$\tilde{\mathbf{y}}_t = \mathbf{\Phi}_1 \tilde{\mathbf{y}}_{t-1} + \mathbf{\Phi}_2 \tilde{\mathbf{y}}_{t-2} + \ldots + \mathbf{\Phi}_P \tilde{\mathbf{y}}_{t-P} + \boldsymbol{\epsilon}_t.$$

Using the deviations form h-step ahead forecasts from a VAR(P) can be computed using the recurrence

$$\mathrm{E}_t[\tilde{\mathbf{y}}_{t+h}] = \mathbf{\Phi}_1 \mathrm{E}_t[\tilde{\mathbf{y}}_{t+h-1}] + \mathbf{\Phi}_2 \mathrm{E}_t[\tilde{\mathbf{y}}_{t+h-2}] + \ldots + \mathbf{\Phi}_P \mathrm{E}_t[\tilde{\mathbf{y}}_{t+h-P}].$$

starting at $\mathrm{E}_t[\tilde{\mathbf{y}}_{t+1}]$. Once the forecast of $\mathrm{E}_t[\tilde{\mathbf{y}}_{t+h}]$ has been computed, the h-step ahead forecast of $\mathbf{y}_{t+h}$ is constructed by adding the long run mean, $\mathrm{E}_t[\mathbf{y}_{t+h}] = \boldsymbol{\mu} + \mathrm{E}_t[\tilde{\mathbf{y}}_{t+h}]$.

5.4.1 Example: The interaction of stock and bond returns

When two series are related in time, univariate forecasts may not adequately capture the feedback between the two and are generally misspecified if the series belong in a VAR. To illustrate the differences, recursively estimated 1-step ahead forecasts were produced from the stock-bond VAR,

$$\begin{bmatrix} VWM_t \\ 10YR_t \end{bmatrix} = \begin{bmatrix} 9.733 \\ 1.058 \end{bmatrix} + \begin{bmatrix} 0.097 & 0.301 \\ -0.095 & 0.299 \end{bmatrix} \begin{bmatrix} VWM_{t-1} \\ 10YR_{t-1} \end{bmatrix} + \begin{bmatrix} \epsilon_{1,t} \\ \epsilon_{2,t} \end{bmatrix}$$

and a simple AR(1) for each series. The data set contains a total of 620 observations. Beginning at observation 381 and continuing until observation 620, the models (the VAR and the two ARs)

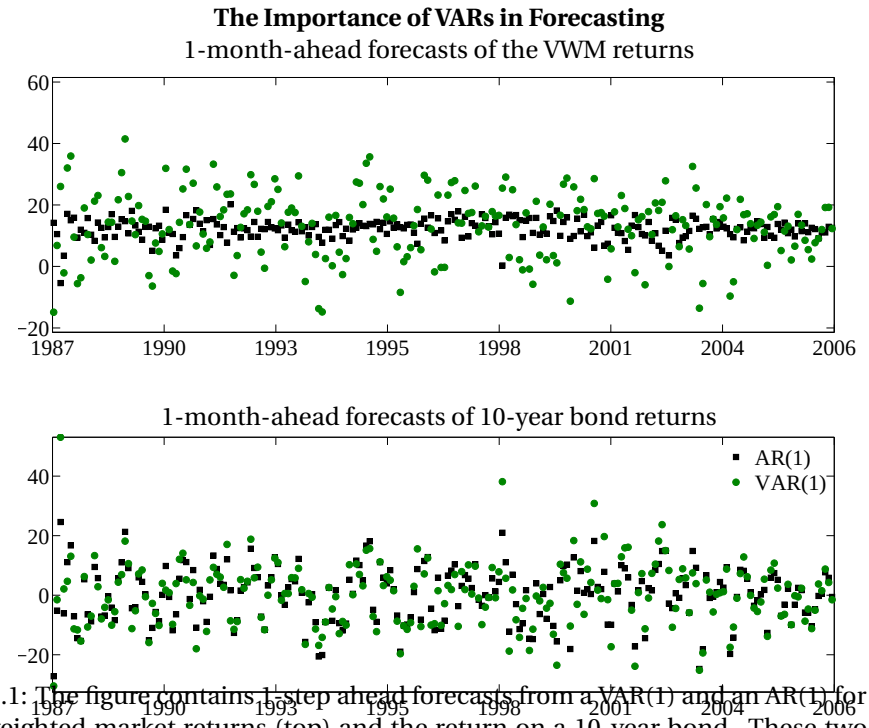

Figure 5.1: The figure contains 1-step ahead forecasts from a VAR(1) and an AR(1) for both the value-weighted market returns (top) and the return on a 10-year bond. These two pictures indicate that the return on the long bond has substantial predictive power for equity returns while the opposite is not true.

---

were estimated using an expanding window of data and 1-step ahead forecasts were computed.[6] Figure 5.1 contains a graphical representation of the differences between the AR(1)s and the VAR(1). The forecasts for the market are substantially different while the forecasts for the 10-year bond return are not. The changes (or lack there of) are simply a function of the model specification: the return on the 10-year bond has predictive power for both series. The VAR(1) is a better model for stock returns (than an AR(1)) although it not meaningfully better for bond returns.

[6] Recursive forecasts computed using an expanding window use data from $t = 1$ to R to estimate any model parameters and to produce a forecast of $R+1$. The sample then grows by one observation and data from 1 to $R + 1$ are used to estimate model parameters and a forecast of $R + 2$ is computed. This pattern continues until the end of the sample. An alternative is to use rolling windows where both the end point and the start point move through time so that the distance between the two is constant.

5.5 Estimation and Identification

Estimation and identification is the first significant break from directly applying the lessons learned from the analogies to univariate modeling to multivariate models. In addition to autocorrelation function and partial autocorrelation functions, vector stochastic processes also have cross-correlation functions (CCFs) and partial cross-correlation functions (PCCFs).

Definition 5.8 (Cross-correlation)**.** The s^{th} cross correlations between two covariance stationary series $\{x_t\}$ and $\{y_t\}$ are defined

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

and

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

where the order of the indices indicates which variable is measured using contemporaneous values and which variable is lagged, $\mathrm{E}[y_t] = \mu_y$ and $\mathrm{E}[x_t] = \mu_x$.

It should be obvious that, unlike autocorrelations, cross-correlations are not symmetric – the order, xy or yx, matters. Partial cross-correlations are defined in a similar manner; the correlation between x_t and y_{t-s} controlling for $y_{t-1}, \ldots, y_{t-(s-1)}$.

Definition 5.9 (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

$$x_t = \phi_0 + \phi_1 y_{t-1} + \ldots + \phi_{s-1} y_{t-(s-1)} + \varphi_{xy,s} y_{t-s} + \epsilon_{x,t} \tag{5.22}$$

and $\varphi_{yx,s}$ in

$$y_t = \phi_0 + \phi_1 x_{t-1} + \ldots + \phi_{s-1} x_{t-(s-1)} + \varphi_{yx,s} x_{t-s} + \epsilon_{x,t} \tag{5.23}$$

where the order of the indices indicates which variable is measured using contemporaneous values and which variable is lagged.

Figure 5.2 contains the CCF (cross-correlation function) and PCCF (partial cross-correlation function) of two first order VARs with identical persistence. The top panel contains the functions for

$$\begin{bmatrix} y_t \\ x_t \end{bmatrix} = \begin{bmatrix} .5 & .4 \\ .4 & .5 \end{bmatrix} \begin{bmatrix} y_{t-1} \\ x_{t-1} \end{bmatrix} + \begin{bmatrix} \epsilon_{1,t} \\ \epsilon_{2,t} \end{bmatrix}$$

while the bottom contains the functions for a trivial VAR

$$\begin{bmatrix} y_t \\ x_t \end{bmatrix} = \begin{bmatrix} .9 & 0 \\ 0 & .9 \end{bmatrix} \begin{bmatrix} y_{t-1} \\ x_{t-1} \end{bmatrix} + \begin{bmatrix} \epsilon_{1,t} \\ \epsilon_{2,t} \end{bmatrix}$$

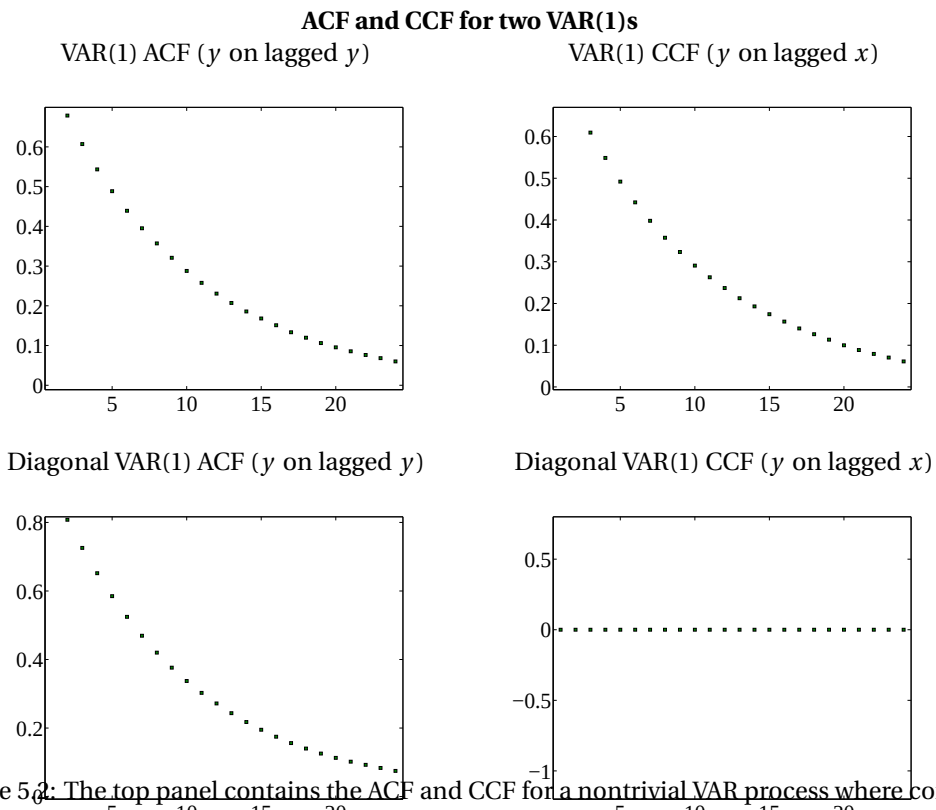

Figure 5.2: The top panel contains the ACF and CCF for a nontrivial VAR process where contemporaneous values depend on both series. The bottom contains the ACF and CCF for a trivial VAR which is simply composed to two AR(1)s.

which is just two AR(1)s in a system. The nontrivial VAR(1) exhibits dependence between both series while the AR-in-disguise shows no dependence between y_t and x_{t-j}, $j > 0$.

With these new tools, it would seem that Box-Jenkins could be directly applied to vector processes, and while technically possible, exploiting the ACF, PACF, CCF and PCCF to determine what type of model is appropriate is difficult. For specifications larger than a bivariate VAR, there are simply too many interactions.

The usual solution is to take a hands off approach as advocated by Sims (1980). The VAR specification should include all variables which theory indicate are relevant and a lag length should be chosen which has a high likelihood of capturing all of the dynamics. Once these values have been set, either a general-to-specific search can be conducted or an information criteria can be used to select the appropriate lag length. In the VAR case, the Akaike IC, Hannan & Quinn (1979) IC and the Schwarz/Bayes IC are given by

$$\begin{aligned}
\text{AIC:} \quad & \ln|\hat{\Sigma}(P)| + k^2 P \frac{2}{T} \\
\text{HQC:} \quad & \ln|\hat{\Sigma}(P)| + k^2 P \frac{2\ln\ln T}{T} \\
\text{SBIC:} \quad & \ln|\hat{\Sigma}(P)| + k^2 P \frac{\ln T}{T}
\end{aligned}$$

where $\hat{\Sigma}(P)$ is the covariance of the residuals using P lags and $|\cdot|$ indicates determinant.[7] The lag length should be chosen to minimize one of these criteria, and the SBIC will always choose a (weakly) smaller model than the HQC which in turn will select a (weakly) smaller model than the AIC. Ivanov & Kilian (2005) recommend the AIC for monthly models and the HQC for quarterly models, unless the sample size is less than 120 quarters in which case the SBIC is preferred. Their recommendation is based on the accuracy of the impulse response function, and so may not be ideal in other applications such as forecasting.

To use a general-to-specific approach, a simple likelihood ratio test can be computed as

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

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 variance is larger which ensures this statistic is positive. The $-P_2 k^2$ term in the log-likelihood is a degree of freedom correction that generally improves small-sample performance. Ivanov & Kilian (2005) recommended against using sequential likelihood ratio testing for selecting lag length.

5.5.1 Example: Campbell's VAR

A lag length selection procedure was conducted using Campbell's VAR. The results are contained in table 5.2. This table contains both the AIC and SBIC values for lags 0 through 12 as well as likelihood ratio test results for testing l lags against $l+1$. Note that the LR and P-val corresponding to lag l is a test of the null of l lags against an alternative of $l+1$ lags. Using the AIC, 12 lags would be selected since it produces the smallest value. If the initial lag length was less than 12, 7 lags would be selected. The HQC and SBIC both choose 3 lags in a specific-to-general search and 12 in a general-to-specific search. Using the likelihood ratio, a general-to-specific procedure chooses 12 lags while a specific-to-general procedure chooses 3. The test statistic for a null of $H_0 : P = 11$ against an alternative that $H_1 : P = 12$ has a p-val of 0.

One final specification search was conducted. Rather than beginning at the largest lag and work down one at a time, a "global search" which evaluates models using every combination of

[7] $\ln|\hat{\Sigma}|$ is, up to an additive constant, the gaussian log-likelihood divided by T, and these 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.

Lag Length	AIC	HQC	BIC	LR	P-val
0	6.39	5.83	5.49	1798	0.00
1	3.28	2.79	2.56	205.3	0.00
2	2.98	2.57	2.45	1396	0.00
3	0.34	0.00	0.00	39.87	0.03
4	0.35	0.08	0.19	29.91	0.23
5	0.37	0.17	0.40	130.7	0.00
6	0.15	0.03	0.37	44.50	0.01
7	0.13	0.08	0.53	19.06	0.79
8	0.17	0.19	0.75	31.03	0.19
9	0.16	0.26	0.94	19.37	0.78
10	0.19	0.36	1.15	27.33	0.34
11	0.19	0.43	1.34	79.26	0.00
12	0.00	0.31	1.33	N/A	N/A

Table 5.2: Normalized values for the AIC and SBIC in Campbell's VAR. The AIC chooses 12 lags while the SBIC chooses only 3. A general-to-specific search would stop at 12 lags since the likelihood ratio test of 12 lags against 11 rejects with a p-value of 0. If the initial number of lags was less than 12, the GtS procedure would choose 6 lags. Note that the LR and P-val corresponding to lag l is a test of the null of l lags against an alternative of $l+1$ lags.

lags up to 12 was computed. This required fitting 4096 VARs which only requires a few seconds on a modern computer.[8] For each possible combination of lags, the AIC and the SBIC were computed. Using this methodology, the AIC search selected lags 1-4, 6, 10 and 12 while the SBIC selected a smaller model with only lags 1, 3 and 12 - the values of these lags indicate that there may be a seasonality in the data. Search procedures of this type are computationally viable for checking up to about 20 lags.

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.

5.6.1 Definition

Granger causality is defined in the negative.

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

Definition 5.10 (Granger causality)**.** A scalar random variable $\{x_t\}$ is said to **not** Granger cause $\{y_t\}$ if
$\mathrm{E}[y_t|x_{t-1}, y_{t-1}, x_{t-2}, y_{t-2}, \ldots] = \mathrm{E}[y_t|, y_{t-1}, y_{t-2}, \ldots]$.[9] That is, $\{x_t\}$ does not Granger cause 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.

$$\begin{bmatrix} x_t \\ y_t \end{bmatrix} = \begin{bmatrix} \phi_{11,1} & \phi_{12,1} \\ \phi_{21,1} & \phi_{22,1} \end{bmatrix} \begin{bmatrix} x_{t-1} \\ y_{t-1} \end{bmatrix} + \begin{bmatrix} \phi_{11,2} & \phi_{12,2} \\ \phi_{21,2} & \phi_{22,2} \end{bmatrix} \begin{bmatrix} x_{t-2} \\ y_{t-2} \end{bmatrix} + \begin{bmatrix} \epsilon_{1,t} \\ \epsilon_{2,t} \end{bmatrix}$$

In this model, if $\phi_{21,1} = \phi_{21,2} = 0$ then $\{x_t\}$ does not Granger cause $\{y_t\}$. If this is the case, it may be tempting to model y_t using

$$y_t = \phi_{22,1} y_{t-1} + \phi_{22,2} y_{t-1} + \epsilon_{2,t}$$

However, it is not; $\epsilon_{1,t}$ and $\epsilon_{2,t}$ can be contemporaneously correlated. If it happens to be the case that $\{x_t\}$ does not Granger cause $\{y_t\}$ and $\epsilon_{1,t}$ and $\epsilon_{2,t}$ have no contemporaneous correlation, then y_t is said to be **weakly exogenous**, and y_t can be modeled completely independently of x_t. Finally it is worth noting that $\{x_t\}$ not Granger causing $\{y_t\}$ says nothing about whether $\{y_t\}$ Granger causes $\{x_t\}$.

One important limitation of GC is that it doesn't account for indirect effects. For example, suppose x_t and y_t are both Granger caused by z_t. When this is the case, x_t will usually Granger cause y_t even when it has no effect once z_t has been conditioned on, and so $E[y_t|y_{t-1}, z_{t-1}, x_{t-1}, \ldots] = E[y_t|y_{t-1}, z_{t-1}, \ldots]$ but $E[y_t|y_{t-1}, x_{t-1}, \ldots] \neq E[y_t|y_{t-1}, \ldots]$.

5.6.2 Testing

Testing for Granger causality in a VAR(P) is usually conducted using likelihood ratio tests. In this specification,

$$\mathbf{y}_t = \mathbf{\Phi}_0 + \mathbf{\Phi}_1 \mathbf{y}_{t-1} + \mathbf{\Phi}_2 \mathbf{y}_{t-2} + \ldots + \mathbf{\Phi}_P \mathbf{y}_{t-P} + \boldsymbol{\epsilon}_t,$$

$\{y_{j,t}\}$ does not Granger cause $\{y_{i,t}\}$ if $\phi_{ij,1} = \phi_{ij,2} = \ldots = \phi_{ij,P} = 0$. The likelihood ratio test can be computed

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

where $\mathbf{\Sigma}_r$ 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 $\mathbf{\Sigma}_u$ is the estimated covariance in the unrestricted VAR(P). If there is no Granger causation in a VAR, it is probably not a good idea to use one.[10]

[9] 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).

[10] 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.

	VWM		LBR		RTB		TERM		DIV	
Exclusion	Stat	P-val	Stat	P-val	Stat	P-val	Stat	P-val	Stat	P-v
VWM	–	–	0.05	0.83	0.06	0.80	2.07	0.15	2.33	0.
LBR	1.54	0.21	–	–	3.64	0.06	0.03	0.86	0.44	0.
RTB	12.05	0.00	2.87	0.09	–	–	49.00	0.00	2.88	0.
TERM	13.55	0.00	5.98	0.01	43.93	0.00	–	–	0.57	0.
DIV	0.16	0.69	0.72	0.40	5.55	0.02	0.06	0.80	–	-
All	5.60	0.23	8.83	0.07	72.29	0.00	56.68	0.00	9.06	0.

Table 5.3: 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.

5.6.3 Example: Campbell's VAR

Campbell's VAR will be used to illustrate testing for Granger causality. Table 5.3 contains the results of Granger causality tests from a VAR which included lags 1, 3 and 12 (as chosen by the "global search" SBIC method) for the 5 series in Campbell's VAR. Tests of a variable causing itself have been omitted as these aren't particularly informative in the multivariate context. The table tests whether the variables in the left hand column Granger cause the variables along the top row. From the table, it can be seen that every variable causes at least one other variable since each row contains a p-val indicating significance using standard test sizes (5 or 10%) since the null is *no* Granger causation. It can also be seen that every variable is caused by another by examining the p-values column by column.

5.7 Impulse Response Function

The second concept new to multivariate time-series analysis is the impulse response function. In the univariate world, the ACF was sufficient to understand how shocks decay. 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 a system which, in turn, can feed back into the original variable. After a few iterations of this cycle, it can be difficult to determine how a shock propagates even in a simple bivariate VAR(1).

5.7.1 Defined

Definition 5.11 (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 $\boldsymbol{\epsilon}$, for any j and i, is defined as the change in y_{it+s}, $s \geq 0$ for a unit shock in $\epsilon_{j,t}$.

This definition is somewhat difficult to parse and the impulse response function can be clearly illustrated through a vector moving average (VMA).[11] As long as $\mathbf{y}_t$ is covariance stationary it must have a VMA representation,

$$\mathbf{y}_t = \boldsymbol{\mu} + \boldsymbol{\epsilon}_t + \boldsymbol{\Xi}_1\boldsymbol{\epsilon}_{t-1} + \boldsymbol{\Xi}_2\boldsymbol{\epsilon}_{t-2} + \ldots$$

Using this VMA, the impulse response y_i with respect to a shock in ϵ_j is simply $\{1, \boldsymbol{\Xi}_{1[ii]}, \boldsymbol{\Xi}_{2[ii]}, \boldsymbol{\Xi}_{3[ii}$ if $i = j$ and $\{0, \boldsymbol{\Xi}_{1[ij]}, \boldsymbol{\Xi}_{2[ij]}, \boldsymbol{\Xi}_{3[ij]}, \ldots\}$ otherwise. The difficult part is computing $\boldsymbol{\Xi}_l$, $l \geq 1$. In the simple VAR(1) model this is easy since

$$\mathbf{y}_t = (\mathbf{I}_k - \boldsymbol{\Phi}_1)^{-1}\boldsymbol{\Phi}_0 + \boldsymbol{\epsilon}_t + \boldsymbol{\Phi}_1\boldsymbol{\epsilon}_{t-1} + \boldsymbol{\Phi}_1^2\boldsymbol{\epsilon}_{t-2} + \ldots.$$

However, in more complicated models, whether higher order VARs or VARMAs, determining the MA(∞) form can be tedious. One surprisingly simply, yet correct, method to compute the elements of $\{\boldsymbol{\Xi}_j\}$ is to simulate the effect of a unit shock of $\epsilon_{j,t}$. Suppose the model is a VAR in deviations form[12],

$$\mathbf{y}_t - \boldsymbol{\mu} = \boldsymbol{\Phi}_1(\mathbf{y}_{t-1} - \boldsymbol{\mu}) + \boldsymbol{\Phi}_2(\mathbf{y}_{t-2} - \boldsymbol{\mu}) + \ldots + \boldsymbol{\Phi}_P(\mathbf{y}_{t-P} - \boldsymbol{\mu}) + \boldsymbol{\epsilon}_t.$$

The impulse responses can be computed by "shocking" $\boldsymbol{\epsilon}_t$ by 1 unit and stepping the process forward. To use this procedure, set $\mathbf{y}_{t-1} = \mathbf{y}_{t-2} = \ldots = \mathbf{y}_{t-P} = \mathbf{0}$ and then begin the simulation by setting $\epsilon_{j,t} = 1$. The 0^{th} impulse will obviously be $\mathbf{e}_j = [\mathbf{0}_{j-1}\, 1\, \mathbf{0}_{k-j}]'$, a vector with a 1 in the j^{th} position and zeros everywhere else. The first impulse will be,

$$\boldsymbol{\Xi}_1 = \boldsymbol{\Phi}_1\mathbf{e}_j,$$

the second will be

$$\boldsymbol{\Xi}_2 = \boldsymbol{\Phi}_1^2\mathbf{e}_j + \boldsymbol{\Phi}_2\mathbf{e}_j$$

and the third is

$$\boldsymbol{\Xi}_3 = \boldsymbol{\Phi}_1^3\mathbf{e}_j + \boldsymbol{\Phi}_1\boldsymbol{\Phi}_2\mathbf{e}_j + \boldsymbol{\Phi}_2\boldsymbol{\Phi}_1\mathbf{e}_j + \boldsymbol{\Phi}_3\mathbf{e}_j.$$

This can be continued to compute any $\boldsymbol{\Xi}_j$.

5.7.2 Correlated Shocks and non-unit Variance

The previous discussion has made use of unit shocks, $\mathbf{e}_j$ which represent a change of 1 in j^{th} error. This presents two problems: actual errors do not have unit variances and are often correlated. The solution to these problems is to use non-standardized residuals and/or correlated residuals.

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

[12] Since the VAR is in deviations form, this formula can be used with any covariance stationary VAR. If the model is not covariance stationary, the impulse response function can still be computed although the unconditional mean must be replaced with a conditional one.

Suppose that the residuals in a VAR have a covariance of $\mathbf{\Sigma}$. To simulate the effect of a shock to element j, $\mathbf{e}_j$ can be replaced with $\tilde{\mathbf{e}}_j = \mathbf{\Sigma}^{1/2}\mathbf{e}_j$ and the impulses can be computed using the procedure previously outlined.

This change has two effects. First, every series will generally have an instantaneous reaction to any shock when the errors are correlated. Second, the choice of matrix square root, $\mathbf{\Sigma}^{1/2}$, matters. There are two matrix square roots: the Choleski and the spectral decomposition. The Choleski square root is a lower triangular matrix which imposes an order to the shocks. Shocking element j (using $\mathbf{e}_j$) has an effect of every series $j,\ldots,k$ but not on $1,\ldots,j-1$. In contrast the spectral matrix square root is symmetric and a shock to the j$^{\text{th}}$ error will generally effect every series instantaneously. Unfortunately there is no right choice. *If* there is a natural ordering in a VAR where shocks to one series can be reasoned to have no contemporaneous effect on the other series, then the Choleski is the correct choice. However, in many situations there is little theoretical guidance and the spectral decomposition is the natural choice.

5.7.3 Example: Impulse Response in Campbell's VAR

Campbell's VAR will be used to illustrate impulse response functions. Figure 5.3 contains the impulse responses of the relative T-bill rate to shocks in the in the four other variables: equity returns, labor income growth, the term premium and the dividend rate. The dotted lines represent 2 standard deviation confidence intervals. The relative T-bill rate increases subsequent to positive shocks in any variable which indicates that the economy is improving and there are inflationary pressures driving up the short end of the yield curve.

5.7.4 Confidence Intervals

Impulse response functions, like the parameters of the VAR, are estimated quantities and subject to statistical variation. Confidence bands can be constructed to determine whether an impulse response is large in a statistically meaningful sense. Since the parameters of the VAR are asymptotically normal (as long as it is stationary and the innovations are white noise), the impulse responses will also be asymptotically normal by applying the δ-method. The derivation of the covariance of the impulse response function is tedious and has no intuitive value. Interested readers can refer to 11.7 in Hamilton (1994). Instead, two computational methods to construct confidence bands for impulse response functions will be described: Monte Carlo and using a procedure known as the *bootstrap*.

5.7.4.1 Monte Carlo Confidence Intervals

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

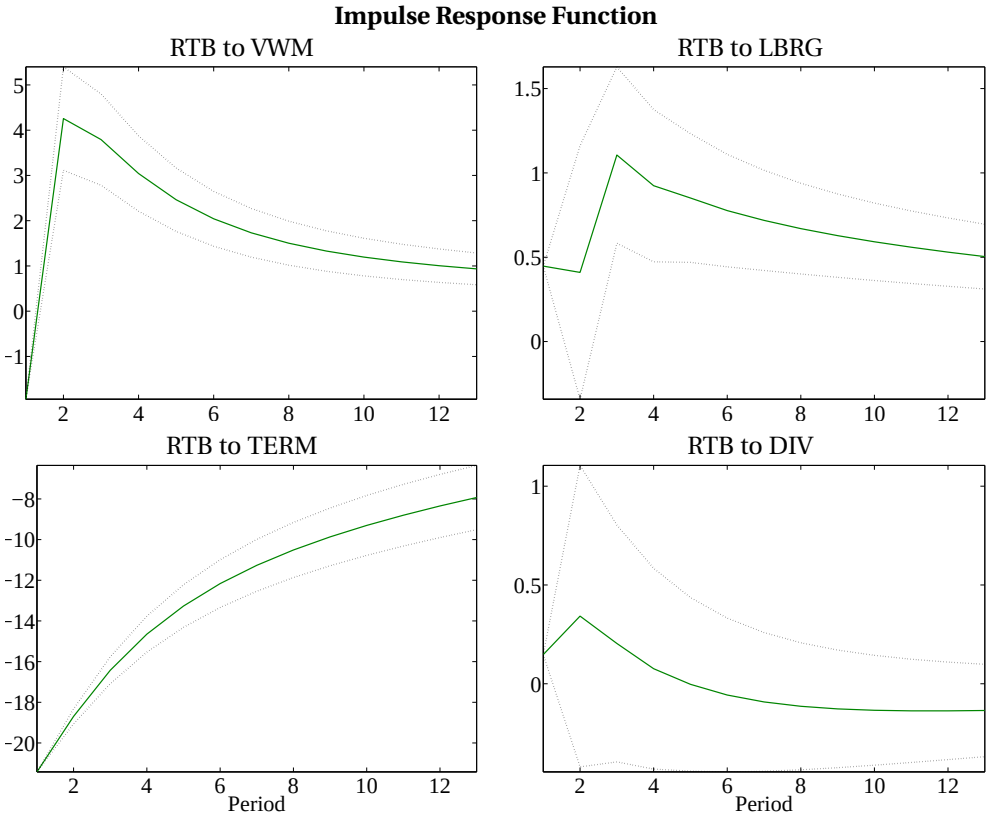

Figure 5.3: Impulse response functions for 12 steps of the response of the relative T-bill rate to equity returns, labor income growth, the term premium rate and the dividend yield. The dotted lines represent 2 standard deviation (in each direction) confidence intervals. All values have been scaled by 1,000.

---

1. Compute $\hat{\mathbf{\Phi}}$ from the initial data and estimate the covariance matrix $\hat{\mathbf{\Lambda}}$ in the asymptotic distribution $\sqrt{T}(\hat{\mathbf{\Phi}} - \mathbf{\Phi}_0) \overset{A}{\sim} N(0, \mathbf{\Lambda})$.[13]

2. Using $\hat{\mathbf{\Phi}}$ and $\hat{\mathbf{\Lambda}}$, generate simulated values $\tilde{\mathbf{\Phi}}_b$ from the asymptotic distribution as $\hat{\mathbf{\Lambda}}^{1/2}\boldsymbol{\epsilon} + \hat{\mathbf{\Phi}}$ where $\boldsymbol{\epsilon} \overset{\text{i.i.d.}}{\sim} N(\mathbf{0}, \mathbf{I})$. These are i.i.d. draws from a $N(\hat{\mathbf{\Phi}}, \hat{\mathbf{\Lambda}})$ distribution.

3. Using $\tilde{\mathbf{\Phi}}_b$, compute the impulse responses $\{\hat{\mathbf{\Xi}}_{j,b}\}$ where $b = 1, 2, \ldots, B$. Save these values.

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

[13] This is an abuse of notation. $\mathbf{\Phi}$ is a matrix and the vec operator is needed to transform it into a vector. Interested readers should see 11.7 in Hamilton (1994) for details on the correct form.

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

The second Monte Carlo method differs only in the method used to compute $\tilde{\mathbf{\Phi}}_b$.

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

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

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

   where $\boldsymbol{\epsilon} \overset{\text{i.i.d.}}{\sim} N(\mathbf{0}, \mathbf{I}_k)$ are i.i.d. multivariate standard normally distributed.

3. Using $\{\tilde{\mathbf{y}}_t\}$, estimate $\tilde{\mathbf{\Phi}}_b$ using a VAR.

4. Using $\tilde{\mathbf{\Phi}}_b$, compute the impulse responses $\{\tilde{\Xi}_{j,b}\}$ where $b = 1, 2, \ldots, B$. 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 as the assumption of i.i.d. normal errors in the latter may be unrealistic. This is particularly true for financial data. The final method, which uses a procedure known as the *bootstrap*, combines the ease of the second with the robustness of the first.

5.7.4.2 Bootstrap Confidence Intervals

The bootstrap is a computational tool which has become popular in recent years primarily due to the significant increase in the computing power of typical PCs. Its name is derived from the expression "pulling oneself up by the bootstraps", a seemingly impossible feat. The idea 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 $\hat{\mathbf{\Phi}}$ from the initial data and estimate the residuals $\hat{\boldsymbol{\epsilon}}_t$.

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

$$\{\hat{\boldsymbol{\epsilon}}_{u_1}, \hat{\boldsymbol{\epsilon}}_{u_2}, \ldots, \hat{\boldsymbol{\epsilon}}_{u_T}\}$$

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

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

4. Using $\{\tilde{\mathbf{y}}_t\}$, compute estimates of $\check{\mathbf{\Phi}}_b$ from a VAR.

5. Using $\check{\mathbf{\Phi}}_b$, compute the impulse responses $\{\check{\Xi}_{j,b}\}$ where $b = 1, 2, \ldots, B$. 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 5^{th} and 95^{th} percentile of this distribution are the confidence intervals.

The bootstrap has many uses in econometrics. Interested readers can find more applications in Efron & Tibshirani (1998).

5.8 Cointegration

Many economic time-series have two properties that make standard VAR analysis unsuitable: they contain one or more unit roots and most equilibrium models specify that deviations between key variables, either in levels or ratios, are transitory. Before formally defining cointegration, consider the case where two important economic variables that contain unit roots, consumption and income, had no long-run relationship. If this were true, the values of these variables would grow arbitrarily far apart given enough time. Clearly this is unlikely to occur and so there must be some long-run relationship between these two time-series. Alternatively, consider the relationship between the spot and future price of oil. Standard finance theory dictates that the future's price, f_t, is a conditionally unbiased estimate of the spot price in period $t+1$, s_{t+1} ($\mathrm{E}_t[s_{t+1}] = f_t$). Additionally, today's spot price is also an unbiased estimate of tomorrow's spot price ($\mathrm{E}_t[s_{t+1}] = s_t$). However, both of these price series 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]

It is also important to note how cointegration is different from stationary VAR analysis. 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 will revert to its unconditional

[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.

[15] This assumes the horizon is short.

mean. In a VAR, both the individual series and linear combinations of the series are stationary. The behavior of cointegrated processes is meaningfully different. Treated in isolation, each process contains a unit root and has shocks with permanent impact. However, when combined with another series, a cointegrated pair will show a tendency to revert towards one another. In other words, a cointegrated pair is mean reverting to a stochastic trend.

Cointegration and **error correction** provide the tools to analyze temporary deviations from long-run equilibria. In a nutshell, cointegrated time-series may show temporary deviations from a long-run trend but are ultimately mean reverting *to this trend*. It may also be useful to relate cointegration to what has been studied thus far: cointegration is to VARs as unit roots are to stationary time-series.

5.8.1 Definition

Recall that an integrated process is defined as a process which is not stationary in levels but is stationary in differences. When this is the case, y_t is said to be 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 I(1) series into I(0) series.

Definition 5.12 (Bivariate Cointegration)**.** Let $\{x_t\}$ and $\{y_t\}$ be two I(1) series. These series are said to be **cointegrated** if there exists a vector $\boldsymbol{\beta}$ with both elements non-zero such that

$$\boldsymbol{\beta}'[x_t \;\; y_t]' = \beta_1 x_t - \beta_2 y_t \sim I(0) \tag{5.24}$$

Put another way, there exists a nontrivial linear combination of x_t and y_t which is stationary. This feature, when present, is a powerful link in the analysis of nonstationary data. When treated individually, the data are extremely persistent; however there is a combination of the data which *is* well behaved. Moreover, in many cases this relationship takes a meaningful form such as $y_t - x_t$. Note that cointegrating relationships are only defined up to a 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 chose one variable to normalize the vector. For example, if $\beta_1 x_t - \beta_2 y_t$ was a cointegrating relationship, one normalized version would be $x_t - \beta_2/\beta_1 y_t = x_t - \tilde{\beta} y_t$.

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

Definition 5.13 (Cointegration)**.** A set of k variables $\mathbf{y}_t$ are said to be cointegrated if at least 2 series are $I(1)$ and there exists a non-zero, reduced rank k by k matrix π such that

$$\pi \mathbf{y}_t \sim I(0). \tag{5.25}$$

The non-zero requirement is obvious: if $\pi = \mathbf{0}$ then $\pi \mathbf{y}_t = \mathbf{0}$ and 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, in order for variables to be *cointegrated* they must also be *integrated*. Thus, if the matrix is full rank, there is no possibility for the common unit roots to cancel and it must have the same order

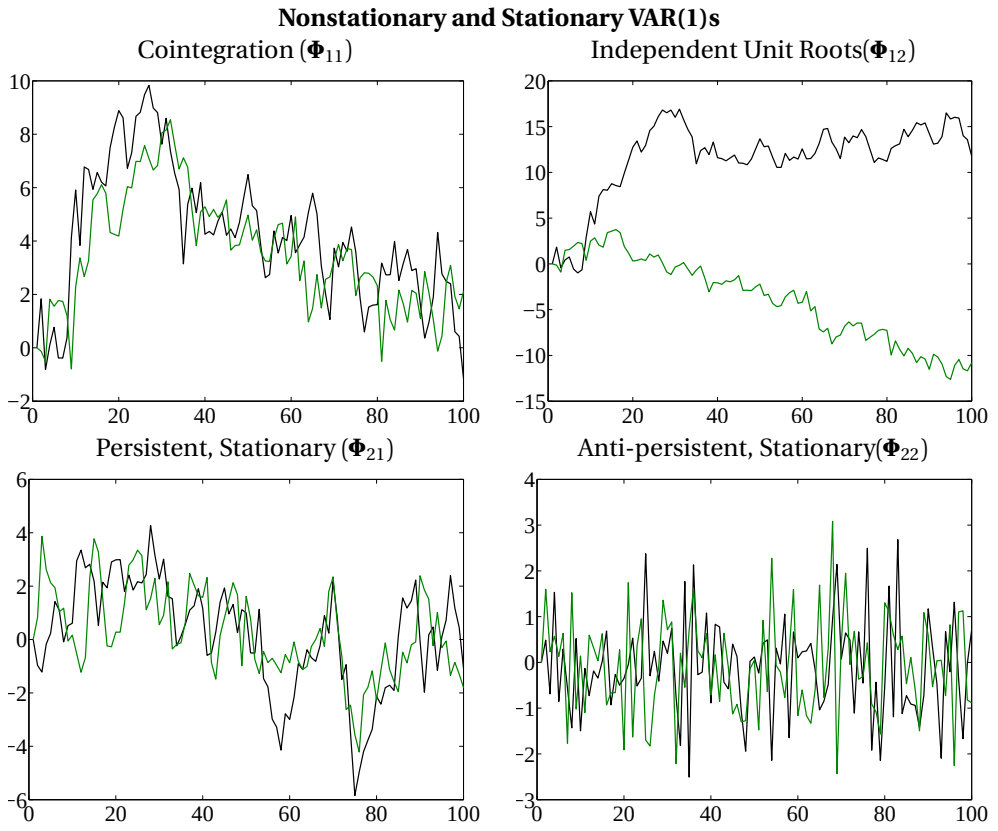

Figure 5.4: A plot of four time-series that all begin at the same point initial value and use the same shocks. All data were generated by $\mathbf{y}_t = \mathbf{\Phi}_{ij}\mathbf{y}_{t-1} + \boldsymbol{\epsilon}_t$ where $\mathbf{\Phi}_{ij}$ varies.

---

of integration before and after the multiplication by π. Finally, the requirement that at least 2 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.

For example, suppose x_t and y_t are cointegrated and $x_t - \beta y_t$ is stationary. One choice for π is

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

To begin developing a feel for cointegration, examine the plots in figure 5.4. These four plots correspond to two nonstationary processes and two stationary processes *all beginning at the same point and all using the same shocks.* These plots contain data from a simulated VAR(1) with different parameters, $\mathbf{\Phi}_{ij}$.

$$\mathbf{y}_t = \mathbf{\Phi}_{ij}\mathbf{y}_{t-1} + \boldsymbol{\epsilon}_t$$

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

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

where λ_i are the eigenvalues of the parameter matrices. Note that the eigenvalues of the nonstationary processes contain the value 1 while the eigenvalues for the stationary processes are all less then 1 (in absolute value). Also, note that the cointegrated process has only one eigenvalue which is unity while the independent unit root process has two. Higher dimension cointegrated systems may contain between 1 and $k-1$ unit eigenvalues. The number of unit eigenvalues indicates the number of unit root "drivers" in a system of equations. The picture presents evidence of another issue in cointegration analysis: it can be very difficult to tell when two series are cointegrated, a feature in common with unit root testing of scalar processes.

5.8.2 Error Correction Models (ECM)

The **Granger representation theorem** provides a key insight to understanding cointegrating relationships. Granger demonstrated that if a system is cointegrated then there exists an error correction model and if there is an error correction model then the system must be cointegrated. The **error correction model** is a form which governs short deviations from the trend (a stochastic trend or unit root). The simplest ECM is given by

$$\begin{bmatrix} \Delta x_t \\ \Delta y_t \end{bmatrix} = \begin{bmatrix} \pi_{11} & \pi_{12} \\ \pi_{21} & \pi_{22} \end{bmatrix} \begin{bmatrix} x_{t-1} \\ y_{t-1} \end{bmatrix} + \begin{bmatrix} \epsilon_{1,t} \\ \epsilon_{2,t} \end{bmatrix} \tag{5.26}$$

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 $x_t - \beta y_t$ is I(0). Substituting this into this equation, equation 5.26 can be rewritten

$$\begin{bmatrix} \Delta x_t \\ \Delta y_t \end{bmatrix} = \begin{bmatrix} \alpha_1 \\ \alpha_2 \end{bmatrix} \begin{bmatrix} 1 & -\beta \end{bmatrix} \begin{bmatrix} x_{t-1} \\ y_{t-1} \end{bmatrix} + \begin{bmatrix} \epsilon_{1,t} \\ \epsilon_{2,t} \end{bmatrix}. \tag{5.27}$$

The short-run dynamics take the forms

$$\Delta x_t = \alpha_1(x_{t-1} - \beta y_{t-1}) + \epsilon_{1,t} \tag{5.28}$$

and

$$\Delta y_t = \alpha_2(x_{t-1} - \beta y_{t-1}) + \epsilon_{2,t}. \tag{5.29}$$

The important elements of this ECM 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. ECMs 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 ECM can be augmented to allow past short-run deviations to also influence present short-run deviations or to include deterministic trends. In vector form, the generalized ECM is

$$\Delta \mathbf{y}_t = \pi_0 + \pi \mathbf{y}_{t-1} + \pi_1 \Delta \mathbf{y}_{t-1} + \pi_2 \Delta \mathbf{y}_{t-2} + \ldots + + \pi_P \Delta \mathbf{y}_{t-P} + \boldsymbol{\epsilon}_t$$

where $\pi \mathbf{y}_{t-1}$ captures the cointegrating relationship, π_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 ECM

It may not be obvious how a cointegrated VAR is transformed into an ECM. Consider a simple cointegrated bivariate VAR(1)

$$\begin{bmatrix} x_t \\ y_t \end{bmatrix} = \begin{bmatrix} .8 & .2 \\ .2 & .8 \end{bmatrix} \begin{bmatrix} x_{t-1} \\ y_{t-1} \end{bmatrix} + \begin{bmatrix} \epsilon_{1,t} \\ \epsilon_{2,t} \end{bmatrix}$$

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

$$\begin{aligned}
\begin{bmatrix} x_t \\ y_t \end{bmatrix} - \begin{bmatrix} x_{t-1} \\ y_{t-1} \end{bmatrix} &= \begin{bmatrix} .8 & .2 \\ .2 & .8 \end{bmatrix} \begin{bmatrix} x_{t-1} \\ y_{t-1} \end{bmatrix} - \begin{bmatrix} x_{t-1} \\ y_{t-1} \end{bmatrix} + \begin{bmatrix} \epsilon_{1,t} \\ \epsilon_{2,t} \end{bmatrix} \\
\begin{bmatrix} \Delta x_t \\ \Delta y_t \end{bmatrix} &= \left(\begin{bmatrix} .8 & .2 \\ .2 & .8 \end{bmatrix} - \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \right) \begin{bmatrix} x_{t-1} \\ y_{t-1} \end{bmatrix} + \begin{bmatrix} \epsilon_{1,t} \\ \epsilon_{2,t} \end{bmatrix} \\
\begin{bmatrix} \Delta x_t \\ \Delta y_t \end{bmatrix} &= \begin{bmatrix} -.2 & .2 \\ .2 & -.2 \end{bmatrix} \begin{bmatrix} x_{t-1} \\ y_{t-1} \end{bmatrix} + \begin{bmatrix} \epsilon_{1,t} \\ \epsilon_{2,t} \end{bmatrix} \\
\begin{bmatrix} \Delta x_t \\ \Delta y_t \end{bmatrix} &= \begin{bmatrix} -.2 \\ .2 \end{bmatrix} \begin{bmatrix} 1 & -1 \end{bmatrix} \begin{bmatrix} x_{t-1} \\ y_{t-1} \end{bmatrix} + \begin{bmatrix} \epsilon_{1,t} \\ \epsilon_{2,t} \end{bmatrix}
\end{aligned} \tag{5.30}$$

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 ECM by recursive substitution. Consider a cointegrated VAR(3),

$$\mathbf{y}_t = \boldsymbol{\Phi}_1 \mathbf{y}_{t-1} + \boldsymbol{\Phi}_2 \mathbf{y}_{t-2} + \boldsymbol{\Phi}_3 \mathbf{y}_{t-3} + \boldsymbol{\epsilon}_t$$

This system will be cointegrated if at least one but fewer than k eigenvalues of $\pi = \boldsymbol{\Phi}_1 + \boldsymbol{\Phi}_2 + \boldsymbol{\Phi}_3 - \mathbf{I}_k$ are not zero. To begin the transformation, add and subtract $\boldsymbol{\Phi}_3 \mathbf{y}_{t-2}$ to the right side

$$\begin{aligned}\mathbf{y}_t &= \mathbf{\Phi}_1\mathbf{y}_{t-1} + \mathbf{\Phi}_2\mathbf{y}_{t-2} + \mathbf{\Phi}_3\mathbf{y}_{t-2} - \mathbf{\Phi}_3\mathbf{y}_{t-2} + \mathbf{\Phi}_3\mathbf{y}_{t-3} + \boldsymbol{\epsilon}_t \\ &= \mathbf{\Phi}_1\mathbf{y}_{t-1} + \mathbf{\Phi}_2\mathbf{y}_{t-2} + \mathbf{\Phi}_3\mathbf{y}_{t-2} - \mathbf{\Phi}_3\Delta\mathbf{y}_{t-2} + \boldsymbol{\epsilon}_t \\ &= \mathbf{\Phi}_1\mathbf{y}_{t-1} + (\mathbf{\Phi}_2 + \mathbf{\Phi}_3)\mathbf{y}_{t-2} - \mathbf{\Phi}_3\Delta\mathbf{y}_{t-2} + \boldsymbol{\epsilon}_t\end{aligned}$$

then add and subtract $(\mathbf{\Phi}_2 + \mathbf{\Phi}_3)\mathbf{y}_{t-1}$ to the right side,

$$\begin{aligned}\mathbf{y}_t &= \mathbf{\Phi}_1\mathbf{y}_{t-1} + (\mathbf{\Phi}_2 + \mathbf{\Phi}_3)\mathbf{y}_{t-1} - (\mathbf{\Phi}_2 + \mathbf{\Phi}_3)\mathbf{y}_{t-1} + (\mathbf{\Phi}_2 + \mathbf{\Phi}_3)\mathbf{y}_{t-2} - \mathbf{\Phi}_3\Delta\mathbf{y}_{t-2} + \boldsymbol{\epsilon}_t \\ &= \mathbf{\Phi}_1\mathbf{y}_{t-1} + (\mathbf{\Phi}_2 + \mathbf{\Phi}_3)\mathbf{y}_{t-1} - (\mathbf{\Phi}_2 + \mathbf{\Phi}_3)\Delta\mathbf{y}_{t-1} - \mathbf{\Phi}_3\Delta\mathbf{y}_{t-2} + \boldsymbol{\epsilon}_t \\ &= (\mathbf{\Phi}_1 + \mathbf{\Phi}_2 + \mathbf{\Phi}_3)\mathbf{y}_{t-1} - (\mathbf{\Phi}_2 + \mathbf{\Phi}_3)\Delta\mathbf{y}_{t-1} - \mathbf{\Phi}_3\Delta\mathbf{y}_{t-2} + \boldsymbol{\epsilon}_t.\end{aligned}$$

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

$$\begin{aligned}\mathbf{y}_t - \mathbf{y}_{t-1} &= (\mathbf{\Phi}_1 + \mathbf{\Phi}_2 + \mathbf{\Phi}_3)\mathbf{y}_{t-1} - \mathbf{y}_{t-1} - (\mathbf{\Phi}_2 + \mathbf{\Phi}_3)\Delta\mathbf{y}_{t-1} - \mathbf{\Phi}_3\Delta\mathbf{y}_{t-2} + \boldsymbol{\epsilon}_t \\ \Delta\mathbf{y}_t &= (\mathbf{\Phi}_1 + \mathbf{\Phi}_2 + \mathbf{\Phi}_3 - \mathbf{I}_k)\mathbf{y}_{t-1} - (\mathbf{\Phi}_2 + \mathbf{\Phi}_3)\Delta\mathbf{y}_{t-1} - \mathbf{\Phi}_3\Delta\mathbf{y}_{t-2} + \boldsymbol{\epsilon}_t.\end{aligned}$$

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

$$\begin{aligned}\mathbf{y}_t - \mathbf{y}_{t-1} &= (\mathbf{\Phi}_1 + \mathbf{\Phi}_2 + \mathbf{\Phi}_3 - \mathbf{I}_k)\mathbf{y}_{t-1} - (\mathbf{\Phi}_2 + \mathbf{\Phi}_3)\Delta\mathbf{y}_{t-1} - \mathbf{\Phi}_3\Delta\mathbf{y}_{t-2} + \boldsymbol{\epsilon}_t \\ \Delta\mathbf{y}_t &= \pi\mathbf{y}_{t-1} + \pi_1\Delta\mathbf{y}_{t-1} + \pi_2\Delta\mathbf{y}_{t-2} + \boldsymbol{\epsilon}_t.\end{aligned} \tag{5.31}$$

which is equivalent to

$$\Delta\mathbf{y}_t = \boldsymbol{\alpha}\boldsymbol{\beta}'\mathbf{y}_{t-1} + \pi_1\Delta\mathbf{y}_{t-1} + \pi_2\Delta\mathbf{y}_{t-2} + \boldsymbol{\epsilon}_t. \tag{5.32}$$

where $\boldsymbol{\alpha}$ contains the speed of adjustment parameters and $\boldsymbol{\beta}$ contains the cointegrating vectors. This recursion can be used to transform any cointegrated VAR(P)

$$\mathbf{y}_{t-1} = \mathbf{\Phi}_1\mathbf{y}_{t-1} + \mathbf{\Phi}_2\mathbf{y}_{t-2} + \ldots + \mathbf{\Phi}_P\mathbf{y}_{t-P} + \boldsymbol{\epsilon}_t$$

into its ECM from

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

using the identities $\pi = -\mathbf{I}_k + \sum_{i=1}^{P}\mathbf{\Phi}_i$ and $\pi_p = -\sum_{i=p+1}^{P}\mathbf{\Phi}_i$.

5.8.2.2 Cointegrating Vectors

The key to understanding cointegration in systems with 3 or more variables is to note that the matrix which governs the cointegrating relationship, π, can always be decomposed into two matrices,

$$\pi = \boldsymbol{\alpha}\boldsymbol{\beta}'$$

where $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ are both k by r matrices where r is the number of cointegrating relationships. For example, suppose the parameter matrix in an ECM was

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

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, $\boldsymbol{\beta}$ can be specified as

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

and $\boldsymbol{\alpha}$ has 6 unknown parameters. $\boldsymbol{\alpha}\boldsymbol{\beta}'$ can be combined 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 $\boldsymbol{\alpha}$ can be trivially solved using the left block of π. Once $\boldsymbol{\alpha}$ 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 solving a trivariate cointegrated VAR for the speed of adjustment coefficients and the cointegrating vectors.

5.8.3 Rank and the number of unit roots

The rank of π is the same as the number of cointegrating vectors since $\pi = \boldsymbol{\alpha}\boldsymbol{\beta}'$ and so if π has rank r, then $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ must both have r linearly independent columns. $\boldsymbol{\alpha}$ contains the speed of adjustment parameters and $\boldsymbol{\beta}$ contains the cointegrating vectors. Note that since there are r cointegrating vectors there are $m = k - r$ distinct unit roots in the system. This relationship holds since when there are k variables, and m distinct unit roots, then there are r distinct linear combinations of the series which will be stationary (except in special circumstances).

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

$$\begin{aligned} y_{1,t} &= \kappa_1 w_{1,t} + \epsilon_{1,t} \\ y_{2,t} &= \kappa_2 w_{1,t} + \epsilon_{2,t} \\ y_{3,t} &= \kappa_3 w_{1,t} + \epsilon_{3,t} \end{aligned}$$

where $\epsilon_{j,t}$ 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 $y_{1,t} - \beta_1 y_{2,t} - \beta_1 y_{3,t}$ it must satisfy

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

to ensure that the unit roots are not present. 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 is a cointegration relationship since

$$y_{1,t} - \frac{\kappa_1}{\kappa_3} y_{3,t} = \kappa_1 w_{1,t} + \epsilon_{1,t} - \frac{\kappa_1}{\kappa_3}\kappa_3 w_{1,t} - \frac{\kappa_1}{\kappa_3}\epsilon_{3,t} = \epsilon_{1,t} - \frac{\kappa_1}{\kappa_3}\epsilon_{3,t}$$

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

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

which again is not identified since there are 2 unknowns and 1 equation. To solve assume $\beta_1 = 0$ and so the solution is $\beta_2 = \kappa 2/\kappa 3$, which is a cointegrating relationship since

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

These solutions are the only two needed since any other definition of the equilibrium would be a linear combination of these two. For example, suppose you choose next to try and normalize on $y_{1,t}$ to define an equilibrium of the form $y_{1,t} - \beta_1 y_{2,t} - \beta_2 y_{3,t}$, and impose that $\beta_3 = 0$ to solve so that $\beta_1 = \kappa_1/\kappa_2$ to produce 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 1 common unit root the set of cointegrating vectors can be expressed as

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

since with only 1 unit root and three series, there are two non-redundant linear combinations of the underlying variables which will be stationary.

Next consider a trivariate system driven by two unit roots,

$$\begin{aligned} y_{1,t} &= \kappa_{11} w_{1,t} + \kappa_{12} w_{2,t} + \epsilon_{1,t} \\ y_{2,t} &= \kappa_{21} w_{1,t} + \kappa_{22} w_{2,t} + \epsilon_{2,t} \\ y_{3,t} &= \kappa_{31} w_{1,t} + \kappa_{32} w_{2,t} + \epsilon_{3,t} \end{aligned}$$

where the errors $\epsilon_{j,t}$ are again covariance stationary. By normalizing the coefficient on $y_{1,t}$ to be 1, it must be the case the weights in the equilibrium condition, $y_{1,t} - \beta_1 y_{2,t} - \beta_2 y_{3,t}$, must satisfy

$$\kappa_{11} = \beta_1 \kappa_{21} + \beta_2 \kappa_{31} \tag{5.33}$$

$$\kappa_{12} = \beta_1 \kappa_{22} + \beta_2 \kappa_{32} \tag{5.34}$$

in order to eliminate both unit roots. This system of 2 equations in 2 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 after the initial normalization and there are no other cointegrating vectors, and so

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

The same line of reasoning extends to k-variate systems driven by m unit roots, and r cointegrating vectors can be constructed by normalizing on the first r elements of y one at a time. In the general case

$$\mathbf{y}_t = \mathbf{K}\mathbf{w}_t + \boldsymbol{\epsilon}_t$$

where $\mathbf{K}$ is a k by m matrix, $\mathbf{w}_t$ a m by 1 set of unit root processes, and $\boldsymbol{\epsilon}_t$ is a k by 1 vector of covariance stationary errors. The cointegrating vectors in this system can be expressed

$$\boldsymbol{\beta} = \begin{bmatrix} \mathbf{I}_r \\ \tilde{\boldsymbol{\beta}} \end{bmatrix} \tag{5.35}$$

where $\mathbf{I}_r$ is an r-dimensional identity matrix and $\tilde{\boldsymbol{\beta}}$ is a m by r matrix of loadings which can be found by solving the set of equations

$$\tilde{\boldsymbol{\beta}} = \mathbf{K}_2^{-1}\mathbf{K}_1' \tag{5.36}$$

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,

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

and in the trivariate system driven by two unit roots,

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

Applying eqs. (5.35) and (5.36) will produce 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 3 equations and only two unknowns. Similarly when $r = k$ there are no unit roots since any linear combination of the series must be stationary.

5.8.3.1 Relationship to Common Features and common trends

Cointegration is special 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 which have been examined are common heteroskedasticity, defined as x_t and y_t are heteroskedastic but there exists a combination, $x_t - \beta y_t$, which is not, and common nonlinearities which are defined in an analogous manner (replacing heteroskedasticity with nonlinearity). When modeling multiple time series, you should always consider whether the aspects you are interested in may be *common*.

5.8.4 Testing

Testing for cointegration shares one important feature with its scalar counterpart (unit root testing): it can be complicated. Two methods will be 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 or the cointegrating relationship 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 more than one cointegrating relationship.

5.8.4.1 Johansen Methodology

The Johansen methodology is the dominant technique to determine whether a system of $I(1)$ variables is cointegrated, and if so, 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.

$$\Delta \mathbf{y}_t = \pi \mathbf{y}_{t-1} + \pi_1 \Delta \mathbf{y}_{t-1} + \ldots + \pi_P \Delta \mathbf{y}_{t-P} \boldsymbol{\epsilon}_t$$

and so the number of non-zero eigenvalues of π is between 1 and $k-1$. If the number of non-zero eigenvalues was k, the system would be stationary. If no non-zero eigenvalues were present, the system would be contain k unit roots. The Johansen framework for cointegration analysis uses the magnitude of these eigenvalues to directly test for cointegration. Additionally, the Johansen methodology 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 $\lambda_1 > \lambda_2 > \ldots > \lambda_k$.[16] The trace statistic is defined

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

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 cannot be rejected. The first trace statistic, $\lambda_{\text{trace}}(0) = -T \sum_{i=1}^{k} \ln(1 - \hat{\lambda}_i)$, tests that null of no cointegrating relationships (e.g. k unit roots) against an alternative that the number of cointegrating relationships is 1 or more. For example, if the were no cointegrating relationships, each of the eigenvalues would be close to zero and $\lambda_{\text{trace}}(0) \approx 0$ since every unit root "driver" corresponds to a zero eigenvalue in π. When the series are cointegrated, π will have one or more non-zero eigenvalues.

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

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

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

defined

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

Intuitively, if there are $r+1$ cointegrating relationships, then the $r+1^{\text{th}}$ ordered eigenvalue should be different from zero and the value of $\lambda_{\max}(r, r+1)$ should be large. On the other hand, if there are only r cointegrating relationships, the $r+1^{\text{th}}$ eigenvalue should be close from zero and the statistic will be small. Again, the distribution is nonstandard but most 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, make note of this and remember to include deterministic terms when estimating the ECM.

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 (General-to-Specific, AIC or SBIC). 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,

$$LR = (T - l \cdot k^2)(\ln |\mathbf{\Sigma}_{l-1}| - \ln |\mathbf{\Sigma}_l|) \sim \chi^2_k.$$

Repeat the test decreasing the number of lags (l) by one each iteration until the LR rejects the null that the smaller model is appropriate.

Step 3: Estimate the selected ECM,

$$\Delta \mathbf{y}_t = \pi \mathbf{y}_{t-1} + \pi_1 \Delta \mathbf{y}_{t-1} + \ldots + \pi_{P-1} \Delta \mathbf{y}_{t-P+1} + \boldsymbol{\epsilon}$$

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

$$\Delta \mathbf{y}_t = \pi_0 + \pi \mathbf{y}_{t-1} + \pi_1 \Delta \mathbf{y}_{t-1} + \ldots + \pi_{P-1} \Delta \mathbf{y}_{t-P+1} + \boldsymbol{\epsilon}$$

should be estimated. Using the λ_{trace} and $\lambda_{\max}$ 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, not excessively heteroskedastic, which will affect 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

		Trace Test		
Null	Alternative	λ_{trace}	Crit. Val.	P-val
$r = 0$	$r \geq 1$	16.77	29.79	0.65
$r = 1$	$r \geq 2$	7.40	15.49	0.53
$r = 2$	$r = 3$	1.86	3.841	0.17
		Max Test		
Null	Alternative	$\lambda_{\max}$	Crit. Val.	P-val
$r = 0$	$r = 1$	9.37	21.13	0.80
$r = 1$	$r = 2$	5.53	14.26	0.67
$r = 2$	$r = 3$	1.86	3.841	0.17

Table 5.4: Results of testing using the Johansen methodology. Unlike the Engle-Granger procedure, no evidence of cointegration is found using either test statistic.

necessary. If the residuals are I(1), the system may not be cointegrated.

5.8.4.2 Example: Consumption Aggregate Wealth

To illustrate cointegration and error correction, three series which have played an important role in the revival of the CCAPM in recent years will be examined. These three series are consumption (c), asset prices (a) and labor income (y). The data were made available by Martin Lettau on his web site,
http://faculty.haas.berkeley.edu/lettau/data_cay.html
and contain quarterly data from 1952:1 until 2009:1.

The Johansen methodology begins by examining the original data for unit roots. Since it has been clearly established that all series have unit roots, this will be skipped. The next step tests eigenvalues of π in the error correction model

$$\Delta \mathbf{y}_t = \pi_0 + \pi \mathbf{y}_{t-1} + \pi_1 \Delta \mathbf{y}_{t-1} + \pi_2 \Delta \mathbf{y}_{t-2} + \ldots + + \pi_P \Delta \mathbf{y}_{t-P} + \boldsymbol{\epsilon}_t.$$

using λ_{trace} and $\lambda_{\max}$ tests. Table 5.4 contains the results of the two tests. These tests are applied sequentially. However, note that all of the p-vals for the null $r = 0$ indicate no significance at conventional levels (5-10%), and so the system appears to contain k unit roots.[17] The Johansen methodology leads to a different conclusion than the Engle-Granger methodology: there is no evidence these three series are cointegrated. This seem counter intuitive, but testing alone cannot provide a reason why this has occurred; only theory can.

[17] Had the first null been rejected, the testing would have continued until a null could not be rejected. The first null not rejected would indicate the cointegrating rank of the system. If all null hypotheses are rejected, then the original system appears stationary, and a further analysis of the I(1) classification of the original data is warranted.

5.8.4.3 Single Cointegrating Vector: Engle-Granger Methodology

The Engle-Granger method exploits the key feature of any cointegrated system where there is a single cointegrating relationship – when data are cointegrated, a linear combination of the series can be constructed that is stationary. If they are not, any linear combination will remain I(1). When there are two variables, the Engle-Granger methodology begins by specifying the cross-section regression

$$y_t = \beta x_t + \epsilon_t$$

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

$$y_t = \beta_1 + \beta x_t + \epsilon_t$$

can be estimated instead if the residuals from the first regression are not mean 0. Once the coefficients have been estimated, the model residuals, $\hat{\epsilon}_t$, can be tested for the presence of a unit root. If x_t and y_t were both I(1) and $\hat{\epsilon}_t$ is I(0), the series are cointegrated. The procedure concludes by using $\hat{\epsilon}_t$ to estimate the error correction model 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. You should 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

$$y_t = \beta_1 + \beta_2 x_t + \epsilon_t$$

using OLS and computing the estimated residuals $\{\hat{\epsilon}_t\}$. Use an ADF test (or DF-GLS for more power) and test $H_0 : \gamma = 0$ against $H_1 : \gamma < 0$ in the regression

$$\Delta\hat{\epsilon}_t = \gamma\hat{\epsilon}_{t-1} + \delta_1\Delta\hat{\epsilon}_{t-1} + \ldots + \delta_p\Delta\hat{\epsilon}_{t-P} + \eta_t.$$

It may be necessary to include deterministic trends. Fortunately, standard procedures for testing time-series for unit roots can be used to examine if this series contains a unit root. If the null is rejected and $\hat{\epsilon}_t$ is stationary, then x_t and y_t appear to be cointegrated. Alternatively, if $\hat{\epsilon}_t$ still contains a unit root, the series are not cointegrated.

Step 3: If a cointegrating relationship is found, specify and estimate the error correction model

$$\begin{bmatrix} \Delta x_t \\ \Delta y_t \end{bmatrix} = \begin{bmatrix} \pi_{01} \\ \pi_{02} \end{bmatrix} + \begin{bmatrix} \alpha_1(y_{t-1} - \beta_1 - \beta_2 x_{t-1}) \\ \alpha_2(y_{t-1} - \beta_1 - \beta_2 x_{t-1}) \end{bmatrix} + \pi_1 \begin{bmatrix} \Delta x_{t-1} \\ \Delta y_{t-1} \end{bmatrix} + \ldots + \pi_P \begin{bmatrix} \Delta x_{t-P} \\ \Delta y_{t-P} \end{bmatrix} + \begin{bmatrix} \eta_{1,t} \\ \eta_{2,t} \end{bmatrix}$$

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 $\hat{\epsilon}_{t-1} = (y_{t-1} - \hat{\beta}_1 - \hat{\beta}_2 x_{t-1})$,

Unit Root Testing on c, a and y			
Series	T-stat	P-val	ADF Lags
c	-1.79	0.39	6
a	-1.20	0.68	3
y	-1.66	0.45	1
$\hat{\epsilon}_t$	-2.91	0.00	2

Table 5.5: Unit root test results. 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. None of these series reject the null of a unit root. The final line contains the results of a unit root test on the estimated residuals where the null is strongly rejected indicating that there *is* a cointegrating relationship between the three. The lags column reports the number of lags used in the ADF procedure as automatically selected using the AIC.

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

and so parameters of these specifications can be estimated using OLS.

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.

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 5.5 and figure 5.5 contain these tests and graphs. The null of a unit root cannot be 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

$$c_t = \beta_1 + \beta_2 a_t + \beta_3 y_t + \epsilon_t$$

and to estimate the long-run relationship using OLS. The estimated cointegrating vector from the first stage OLS was $[1 \quad -0.170 \quad -0.713]$ which corresponds to a long-run relationship of $c_t - 0.994 - 0.170 a_t - 0.713 y_t$. Finally, the residuals were tested for the presence of a unit root. The results of this test are in the final line of table 5.5 and indicate a strong rejection of a unit root in the errors. Based on the Engle-Granger procedure, these three series appear to be cointegrated.

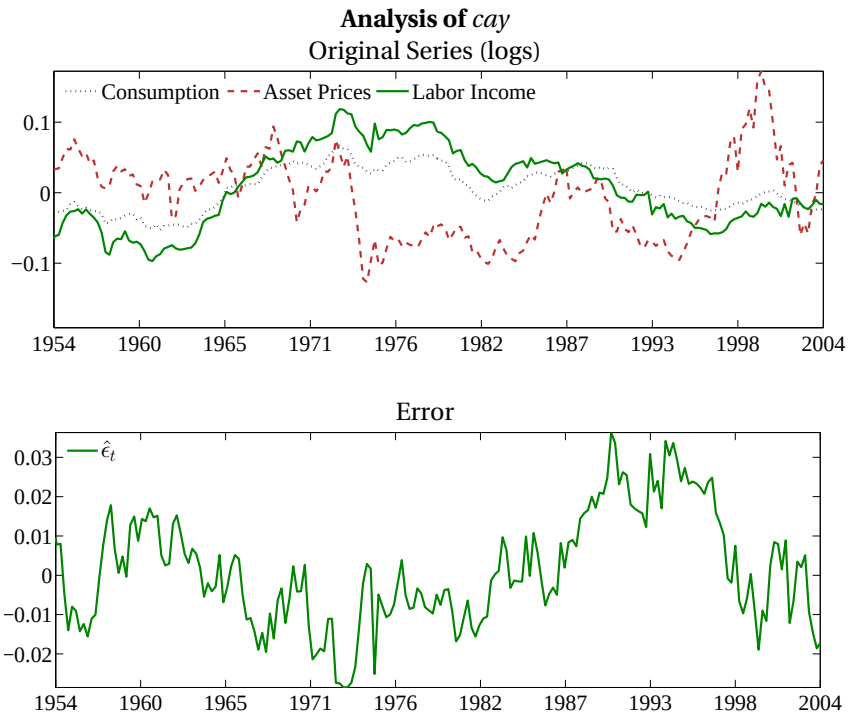

Figure 5.5: The top panel contains plots of time detrended residuals from regressions of consumption, asset prices and labor income on a linear time trend. These series may contain unit roots and are clearly highly persistent. The bottom panel contains a plot of $\epsilon_t = c_t - 0.994 - 0.170a_t - 0.713y_t$ which is commonly known as the cay scaling factor (pronounced consumption-aggregate wealth). The null of a unit root is rejected using the Engle-Granger procedure for this series indicating that the three original series are cointegrated.

5.8.5 Spurious Regression and Balance

When two related I(1) variables are regressed on one another, the cointegrating relationship dominates and the regression coefficients can be directly interpreted as the cointegrating vectors. However, when two unrelated I(1) variables are regressed on one another, the regression coefficient 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

$$x_t = \beta y_t + \epsilon_t$$

$\hat{\beta}$ 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 which indicates 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, use the Engle-Granger procedure as a diagnostic check.

Balance is another important concept when data which contain both stationary and integrated data. An equation is said to be balanced if all variables have the same order of integration. The usual case occurs when a stationary variable (I(0)) is related to other stationary variables. However, other situation may arise and it is useful to consider the four possibilities:

- 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 the examine whether the short-run dynamics in the differenced I(1) 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 meaningful in explaining economic phenomena. Unlike spurious regressions, the t-stat still has a standard asymptotic distribution although caution is needed as small sample properties can be very poor. This is a common problem in finance where a stationary variable, returns on the market, are regressed on a very persistent "predictor" (such as the default premium or dividend yield).

5.9 Cross-sectional Regression with Time-series Data

Finance often requires cross-sectional regressions to be estimated using data that occur sequentially. Chapter 1 used n to index the observations in order to avoid any issues specific to running regressions with time-series data,

$$y_n = \beta_1 x_{n1} + \beta_2 x_{n2} + \ldots + \beta_k x_{nk} + \epsilon_n, \tag{5.37}$$

The observation index above can be replaced with t to indicate that the data used in the regression are from a time-series,

$$y_t = \beta_1 x_{t1} + \beta_2 x_{t2} + \ldots + \beta_k x_{tk} + \epsilon_t. \tag{5.38}$$

Also recall the five assumptions used in establishing the asymptotic distribution of the parameter estimated (recast with time-series indices):

Assumption 5.1 (Linearity)**.** $y_t = \mathbf{x}_t\boldsymbol{\beta} + \epsilon_t$

$\{(\mathbf{x}_t, \epsilon_t)\}$ *is a strictly stationary and ergodic sequence.*

$\mathrm{E}[\mathbf{x}_t'\mathbf{x}_t] = \boldsymbol{\Sigma}_{\mathbf{XX}}$ *is non-singular and finite.*

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

$\mathrm{E}[x_{j,t}^4] < \infty$, $j = 1, 2, \ldots, k$, $t = 1, 2, \ldots$ *and* $\mathrm{E}[\epsilon_t^2] = \sigma^2 < \infty$, $t = 1, 2, \ldots$.

The key assumption often violated in applications using time-series data is assumption **??**, that the scores from the linear regression, $\mathbf{x}_t'\epsilon_t$ are a martingale with respect to the time $t-1$ information set, $\mathcal{F}_{t-1}$. When the scores are not a MDS, it is usually the case that the errors from the model, ϵ_t, can be predicted by variables in $\mathcal{F}_{t-1}$, often their own lagged values. The MDS assumption featured prominently in two theorems about the asymptotic distribution of $\hat{\boldsymbol{\beta}}$ and a consistent estimators of its covariance.

Theorem 5.3. *Under assumptions 3.1 and* **??** - **??**

$$\sqrt{T}(\hat{\boldsymbol{\beta}}_T - \boldsymbol{\beta}) \xrightarrow{d} N(0, \boldsymbol{\Sigma}_{\mathbf{XX}}^{-1}\mathbf{S}\boldsymbol{\Sigma}_{\mathbf{XX}}^{-1}) \tag{5.39}$$

where $\boldsymbol{\Sigma}_{\mathbf{XX}} = \mathrm{E}[\mathbf{x}_t'\mathbf{x}_t]$ *and* $\mathbf{S} = \mathrm{V}[T^{-1/2}\mathbf{X}'\boldsymbol{\epsilon}]$

Under assumptions 3.1 and **??** - **??**,

$$\begin{aligned}
\hat{\boldsymbol{\Sigma}}_{\mathbf{XX}} &= T^{-1}\mathbf{X}'\mathbf{X} \xrightarrow{p} \boldsymbol{\Sigma}_{\mathbf{XX}} \\
\hat{\mathbf{S}} &= T^{-1}\sum_{n=1}^{T} e_t^2\mathbf{x}_t'\mathbf{x}_t \xrightarrow{p} \mathbf{S} \\
&= T^{-1}\left(\mathbf{X}'\hat{\mathbf{E}}\mathbf{X}\right)
\end{aligned}$$

and

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

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

The major change when the assumption of martingale difference scores is relaxed is that a more complicated covariance estimator is required to estimate the variance of $\hat{\boldsymbol{\beta}}$. A modification from White's covariance estimator is needed to pick up the increase in the long run-variance due to the predictability of the scores ($\mathbf{x}_t\epsilon_t$). In essence, the correlation in the scores reduces the amount of "unique" information present in the data. The standard covariance estimator

assumes that the scores are uncorrelated with their past and so each contributes its full share to the precision to $\hat{\boldsymbol{\beta}}$.

Heteroskedasticity Autocorrelation Consistent (HAC) covariance estimators address this issue. Before turning attention to the general case, consider the simple differences that arise in the estimation of the unconditional mean (a regression on a constant) when the errors are a white noise process and when the errors follow a MA(1).

5.9.1 Estimating the mean with time-series errors

To understand why a Newey-West estimator may be needed, consider estimating the mean in two different setups, the first where standard and the shock, $\{\epsilon_t\}$, is assumed to be a white noise process with variance σ^2, and the second where the shock follows an MA(1).

5.9.1.1 White Noise Errors

Define the data generating process for y_t,

$$y_t = \mu + \epsilon_t$$

where $\{\epsilon_t\}$ is a white noise process. It's trivial to show that

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

directly from the white noise assumption. Define the sample mean in the usual way,

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

The sample mean is unbiased,

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

The variance of the mean estimator exploits the white noise property which ensures $\mathrm{E}[\epsilon_i \epsilon_j]$=0 whenever $i \neq j$.

$$
\begin{aligned}
\mathrm{V}[\hat{\mu}] &= \mathrm{E}[(T^{-1}\sum_{t=1}^{T} y_t - \mu)^2] \\
&= \mathrm{E}[(T^{-1}\sum_{t=1}^{T} \epsilon_t)^2] \\
&= \mathrm{E}[T^{-2}(\sum_{t=1}^{T} \epsilon_t^2 + \sum_{r=1}^{T}\sum_{s=1,r\neq s}^{T} \epsilon_r\epsilon_s)] \\
&= T^{-2}\sum_{t=1}^{T} \mathrm{E}[\epsilon_t^2] + T^{-2}\sum_{r=1}^{T}\sum_{s=1,r\neq s}^{T} \mathrm{E}[\epsilon_r\epsilon_s] \\
&= T^{-2}\sum_{t=1}^{T} \sigma^2 + T^{-2}\sum_{r=1}^{T}\sum_{s=1,r\neq s}^{T} 0 \\
&= T^{-2}T\sigma^2 \\
&= \frac{\sigma^2}{T},
\end{aligned}
$$

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

5.9.1.2 MA(1) errors

Consider a modification of the original model where the error process ($\{\eta_t\}$) is a mean zero MA(1) constructed from white noise shocks ($\{\epsilon_t\}$).

$$\eta_t = \theta\epsilon_{t-1} + \epsilon_t$$

The properties of the error can be easily derived. The mean is 0,

$$\mathrm{E}[\eta_t] = \mathrm{E}[\theta\epsilon_{t-1} + \epsilon_t] = \theta\mathrm{E}[\epsilon_{t-1}] + \mathrm{E}[\epsilon_t] = \theta 0 + 0 = 0$$

and the variance depends on the MA parameter,

$$
\begin{aligned}
\mathrm{V}[\eta_t] &= \mathrm{E}[(\theta\epsilon_{t-1} + \epsilon_t)^2] \\
&= \mathrm{E}[\theta^2\epsilon_{t-1}^2 + 2\epsilon_t\epsilon_{t-1} + \epsilon_t^2] \\
&= \mathrm{E}[\theta^2\epsilon_{t-1}^2] + 2\mathrm{E}[\epsilon_t\epsilon_{t-1}] + \mathrm{E}[\epsilon_t^2] \\
&= \theta^2\sigma^2 + 2\cdot 0 + \sigma^2 \\
&= \sigma^2(1+\theta^2).
\end{aligned}
$$

The DGP for y_t has the same form,

$$y_t = \mu + \eta_t$$

and the mean and variance of y_t are

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

These follow from the derivations in chapter 4 for the MA(1) model. More importantly, the usual mean estimator is unbiased.

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

$$\begin{aligned}
\mathrm{E}[\hat{\mu}] &= \mathrm{E}[T^{-1}\sum_{t=1}^{T} y_t] \\
&= T^{-1}\sum_{t=1}^{T} \mathrm{E}[y_t] \\
&= T^{-1}\sum_{t=1}^{T} \mu \\
&= \mu,
\end{aligned}$$

although its variance is *different*. The difference is due to the fact that η_t is autocorrelated and so $\mathrm{E}[\eta_t\eta_{t-1}] \neq 0$.

$$\begin{aligned}
\mathrm{V}[\hat{\mu}] &= \mathrm{E}[(T^{-1}\sum_{t=1}^{T} y_t - \mu)^2] \\
&= \mathrm{E}[(T^{-1}\sum_{t=1}^{T} \eta_t)^2] \\
&= \mathrm{E}[T^{-2}(\sum_{t=1}^{T}\eta_t^2 + 2\sum_{t=1}^{T-1}\eta_t\eta_{t+1} + 2\sum_{t=1}^{T-2}\eta_t\eta_{t+2} + \ldots + 2\sum_{t=1}^{2}\eta_t\eta_{t+T-2} + 2\sum_{t=1}^{1}\eta_t\eta_{t+T-1})] \\
&= T^{-2}\sum_{t=1}^{T}\mathrm{E}[\eta_t^2] + 2T^{-2}\sum_{t=1}^{T-1}\mathrm{E}[\eta_t\eta_{t+1}] + 2T^{-2}\sum_{t=1}^{T-2}\mathrm{E}[\eta_t\eta_{t+2}] + \ldots + \\
&\qquad 2T^{-2}\sum_{t=1}^{2}\mathrm{E}[\eta_t\eta_{t+T-2}] + 2T^{-2}\sum_{t=1}^{1}\mathrm{E}[\eta_t\eta_{t+T-1}]
\end{aligned}$$

$$= T^{-2}\sum_{t=1}^{T}\gamma_0 + 2T^{-2}\sum_{t=1}^{T-1}\gamma_1 + 2T^{-2}\sum_{t=1}^{T-2}\gamma_2 + \ldots + 2T^{-2}\sum_{t=1}^{2}\gamma_{T-2} + 2T^{-2}\sum_{t=1}^{1}\gamma_{T-1}$$

where $\gamma_0 = \mathrm{E}[\eta_t^2] = \mathrm{V}[\eta_t]$ and $\gamma_s = \mathrm{E}[\eta_t\eta_{t-s}]$. The two which are non-zero in this specification are γ_0 and γ_1.

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

since $\gamma_s = 0$, $s \geq 2$ in a MA(1). Returning to the variance of $\hat{\mu}$,

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

and so when the errors are autocorrelated, the usual mean estimator will have a different variance, one which reflects the dependence in the errors, and so it is **not** that case that

$$\mathrm{V}[\hat{\mu}] = \frac{\gamma_0}{T}.$$

This simple illustration captures the basic idea behind the Newey-West covariance estimator, which is defined,

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

When $L = 1$, the only weight is $2(1 - \frac{1}{2}) = 2\frac{1}{2}$ 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 2 since $\lim_{L\to\infty} 1 - \frac{1}{L+1} = 1$ and the Newey-West covariance estimator will, asymptotically, include the important terms from the covariance, $\gamma_0 + 2\gamma_1$, with the correct weights. What happens when we use $\hat{\sigma}_{NW}^2$ instead of the usual variance estimator? As L grows large,

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

and the variance of the estimated mean can be estimated using σ^2_{NW},

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

As a general principle, *the variance of the sum is not the sum of the variances* – this statement is only true when the errors are uncorrelated. Using a HAC covariance estimator allows for time-series dependence and leads to correct inference as long as L grows with the sample size.[18]

It is tempting to estimate $\hat{\gamma}_0$ and $\hat{\gamma}_1$ and use the natural estimator $\hat{\sigma}^2_{HAC} = \hat{\gamma}_0 + 2\hat{\gamma}_1$? Unfortunately this estimator is not guaranteed to be positive, while the Newey-West estimator, $\gamma_0 + \gamma_1$ (when L=1) is always (weakly) positive. More generally, for any choice of L, the Newey-West covariance estimator, $\hat{\sigma}^2_{NW}$, is guaranteed to be positive while the unweighted estimator, $\hat{\sigma}^2_{HAC} = \hat{\gamma}_0 + 2\hat{\gamma}_1 + 2\hat{\gamma}_2 + \ldots + 2\hat{\gamma}_L$, is not. This ensures that the variance estimator passes a minimal sanity check.

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

There are two solutions to working with cross-section data that have autocorrelated errors. The direct method is to change a cross-sectional model in to a time-series model which includes both contemporaneous effects of $\mathbf{x}_t$ as well as lagged values of y_t and possibly $\mathbf{x}_t$. This approach will need to include enough lags so that the errors are white noise. If this can be accomplished, White's heteroskedasticity (but not autocorrelation) consistent covariance estimator can be used and the problem has been solved. The second approach modifies the covariance estimator to explicitly capture the dependence in the data.

The key to 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 it 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). A HAC estimator for a linear regression can be defined as an estimator of the form

$$\hat{\mathbf{S}}_{NW} = T^{-1} \left(\sum_{t=1}^{T} e_t^2 \mathbf{x}_t' \mathbf{x}_t + \sum_{l=1}^{L} w_l \left(\sum_{s=l+1}^{T} e_s e_{s-l} \mathbf{x}_s' \mathbf{x}_{s-l} + \sum_{q=l+1}^{T} e_{q-l} e_q \mathbf{x}_{q-l}' \mathbf{x}_q \right) \right) \tag{5.40}$$

[18] Allowing L to grow at the rate $T^{\frac{1}{3}}$ has been shown to be optimal in a certain sense related to testing.

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

where $\{w_l\}$ are a set of weights. The Newey-West estimator is produced by setting $w_l = 1 - \frac{l}{L+1}$. Other estimators can be computed using different weighting schemes.

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
- Decompose the π matrix in to $\boldsymbol{\alpha}$, the adjustment coefficient, and $\boldsymbol{\beta}$, the cointegrating vectors.

5.A.1 Stationary VAR

Consider the VAR(1):

$$\begin{bmatrix} x_t \\ y_t \\ z_t \end{bmatrix} = \begin{bmatrix} .9 & -.4 & .2 \\ .2 & .8 & -.3 \\ .5 & .2 & .1 \end{bmatrix} \begin{bmatrix} x_{t-1} \\ y_{t-1} \\ z_{t-1} \end{bmatrix} + \begin{bmatrix} \epsilon_{1,t} \\ \epsilon_{2,t} \\ \epsilon_{3,t} \end{bmatrix}$$

Easy method to determine the stationarity of this VAR is to compute the eigenvalues of the parameter matrix. If the eigenvalues are all less than one in modulus, the VAR(1) is stationary. These values are 0.97, 0.62, and 0.2. Since these are all less then one, the model is stationary. An alternative method is to transform the model into an ECM

$$\begin{bmatrix} \Delta x_t \\ \Delta y_t \\ \Delta z_t \end{bmatrix} = \left(\begin{bmatrix} .9 & -.4 & .2 \\ .2 & .8 & -.3 \\ .5 & .2 & .1 \end{bmatrix} - \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \right) \begin{bmatrix} x_{t-1} \\ y_{t-1} \\ z_{t-1} \end{bmatrix} + \begin{bmatrix} \epsilon_{1,t} \\ \epsilon_{2,t} \\ \epsilon_{3,t} \end{bmatrix}$$
$$\begin{bmatrix} \Delta x_t \\ \Delta y_t \\ \Delta z_t \end{bmatrix} = \begin{bmatrix} -.1 & -.4 & .2 \\ .2 & -.2 & -.3 \\ .5 & .2 & -.9 \end{bmatrix} \begin{bmatrix} x_{t-1} \\ y_{t-1} \\ z_{t-1} \end{bmatrix} + \begin{bmatrix} \epsilon_{1,t} \\ \epsilon_{2,t} \\ \epsilon_{3,t} \end{bmatrix}$$
$$\Delta \mathbf{w}_t = \pi \mathbf{w}_t + \boldsymbol{\epsilon}_t$$

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.

5.A.2 Independent Unit Roots

This example is trivial,

$$\begin{bmatrix} x_t \\ y_t \\ z_t \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x_{t-1} \\ y_{t-1} \\ z_{t-1} \end{bmatrix} + \begin{bmatrix} \epsilon_{1,t} \\ \epsilon_{2,t} \\ \epsilon_{3,t} \end{bmatrix}$$

The eigenvalues are trivially 1 and the ECM is given by

$$\begin{bmatrix} \Delta x_t \\ \Delta y_t \\ \Delta z_t \end{bmatrix} = \left(\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} - \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \right) \begin{bmatrix} x_{t-1} \\ y_{t-1} \\ z_{t-1} \end{bmatrix} + \begin{bmatrix} \epsilon_{1,t} \\ \epsilon_{2,t} \\ \epsilon_{3,t} \end{bmatrix}$$

$$\begin{bmatrix} \Delta x_t \\ \Delta y_t \\ \Delta z_t \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} x_{t-1} \\ y_{t-1} \\ z_{t-1} \end{bmatrix} + \begin{bmatrix} \epsilon_{1,t} \\ \epsilon_{2,t} \\ \epsilon_{3,t} \end{bmatrix}$$

and the rank of π is obviously 0, so these are three independent unit roots.

5.A.3 Cointegrated with 1 cointegrating relationship

Consider the VAR(1):

$$\begin{bmatrix} x_t \\ y_t \\ z_t \end{bmatrix} = \begin{bmatrix} 0.8 & 0.1 & 0.1 \\ -0.16 & 1.08 & 0.08 \\ 0.36 & -0.18 & 0.82 \end{bmatrix} \begin{bmatrix} x_{t-1} \\ y_{t-1} \\ z_{t-1} \end{bmatrix} + \begin{bmatrix} \epsilon_{1,t} \\ \epsilon_{2,t} \\ \epsilon_{3,t} \end{bmatrix}$$

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

$$\begin{bmatrix} \Delta x_t \\ \Delta y_t \\ \Delta z_t \end{bmatrix} = \left(\begin{bmatrix} 0.8 & 0.1 & 0.1 \\ -0.16 & 1.08 & 0.08 \\ 0.36 & -0.18 & 0.82 \end{bmatrix} - \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \right) \begin{bmatrix} x_{t-1} \\ y_{t-1} \\ z_{t-1} \end{bmatrix} + \begin{bmatrix} \epsilon_{1,t} \\ \epsilon_{2,t} \\ \epsilon_{3,t} \end{bmatrix}$$

$$\begin{bmatrix} \Delta x_t \\ \Delta y_t \\ \Delta z_t \end{bmatrix} = \begin{bmatrix} -0.2 & 0.1 & 0.1 \\ -0.16 & 0.08 & 0.08 \\ 0.36 & -0.18 & -0.18 \end{bmatrix} \begin{bmatrix} x_{t-1} \\ y_{t-1} \\ z_{t-1} \end{bmatrix} + \begin{bmatrix} \epsilon_{1,t} \\ \epsilon_{2,t} \\ \epsilon_{3,t} \end{bmatrix}$$

and the eigenvalues of π are 0, 0 and -.3 indicating it has rank one. Remember, in a cointegrated system, the number of cointegrating vectors is the rank of π. In this example, there is one cointegrating vector, which can be solved for by transforming π into row-echelon form,

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

So $\boldsymbol{\beta} = [1 \quad -0.5 \quad -0.5]'$ and $\boldsymbol{\alpha}$ can be solved for by noting that

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

and so $\boldsymbol{\alpha} = [-.2 \quad -.16 \quad 0.36]'$ is the first column of π.

5.A.4 Cointegrated with 2 cointegrating relationships

Consider the VAR(1):

$$\begin{bmatrix} x_t \\ y_t \\ z_t \end{bmatrix} = \begin{bmatrix} 0.3 & 0.4 & 0.3 \\ 0.1 & 0.5 & 0.4 \\ 0.2 & 0.2 & 0.6 \end{bmatrix} \begin{bmatrix} x_{t-1} \\ y_{t-1} \\ z_{t-1} \end{bmatrix} + \begin{bmatrix} \epsilon_{1,t} \\ \epsilon_{2,t} \\ \epsilon_{3,t} \end{bmatrix}$$

the eigenvalues of the parameter matrix are 1, .2+.1i and .2-.1i. The ECM form of this model is

$$\begin{bmatrix} \Delta x_t \\ \Delta y_t \\ \Delta z_t \end{bmatrix} = \left(\begin{bmatrix} 0.3 & 0.4 & 0.3 \\ 0.1 & 0.5 & 0.4 \\ 0.2 & 0.2 & 0.6 \end{bmatrix} - \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \right) \begin{bmatrix} x_{t-1} \\ y_{t-1} \\ z_{t-1} \end{bmatrix} + \begin{bmatrix} \epsilon_{1,t} \\ \epsilon_{2,t} \\ \epsilon_{3,t} \end{bmatrix}$$

$$\begin{bmatrix} \Delta x_t \\ \Delta y_t \\ \Delta z_t \end{bmatrix} = \begin{bmatrix} -0.7 & 0.4 & 0.3 \\ 0.1 & -0.5 & 0.4 \\ 0.2 & 0.2 & -0.4 \end{bmatrix} \begin{bmatrix} x_{t-1} \\ y_{t-1} \\ z_{t-1} \end{bmatrix} + \begin{bmatrix} \epsilon_{1,t} \\ \epsilon_{2,t} \\ \epsilon_{3,t} \end{bmatrix}$$

and the eigenvalues of π are 0, $-0.8+0.1i$ and $-0.8-0.1i$ indicating it has rank two (note that two of the eigenvalues are complex). Remember, in a cointegrated system, the number of cointegrating vectors is the rank of π. In this example, there are two cointegrating vectors, which can be solved for by transforming π into row-echelon form,

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

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

$\boldsymbol{\beta}$ is the transpose of first two rows of the row-echelon form,

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

and $\boldsymbol{\alpha}$ can be solved for by noting that

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

and so

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

is the first two columns of π.

Exercises

Shorter Questions

Problem 5.1. Under what conditions are 2 random variables cointegrated?

Problem 5.2. Suppose $\mathbf{y}_t = \mathbf{\Phi}_0 + \mathbf{\Phi}_1 \mathbf{y}_{t-1} + \boldsymbol{\epsilon}_t$ where $\mathbf{y}_t$ is a K by 1 vector values variable and $\mathbf{\Phi}_0$ and $\mathbf{\Phi}_1$ 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 $\epsilon_t = \nu_t + \theta_1 \nu_{t-1}$ where ν_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.

$$
\begin{aligned}
x_t &= \eta_{1,t} + \theta_{11}\eta_{1,t-1} + \theta_{12}\eta_{2,t-1} \\
y_t &= \eta_{2,t} + \theta_{21}\eta_{1,t-1} + \theta_{22}\eta_{2,t-1}
\end{aligned}
$$

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 $\mathbf{\Gamma}_j$ 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

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

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

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

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)

$$\begin{bmatrix} x_t \\ y_t \end{bmatrix} = \begin{bmatrix} 1.4 & .4 \\ -.6 & .4 \end{bmatrix} \begin{bmatrix} x_{t-1} \\ y_{t-1} \end{bmatrix} + \begin{bmatrix} \epsilon_{1t} \\ \epsilon_{2t} \end{bmatrix}$$

(b)

$$\begin{bmatrix} x_t \\ y_t \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} x_{t-1} \\ y_{t-1} \end{bmatrix} + \begin{bmatrix} \epsilon_{1t} \\ \epsilon_{2t} \end{bmatrix}$$

(c)

$$\begin{bmatrix} x_t \\ y_t \end{bmatrix} = \begin{bmatrix} .8 & 0 \\ .2 & .4 \end{bmatrix} \begin{bmatrix} x_{t-1} \\ y_{t-1} \end{bmatrix} + \begin{bmatrix} \epsilon_{1t} \\ \epsilon_{2t} \end{bmatrix}$$

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

$$y_t = \beta_1 + \beta_2 x_t + \epsilon_t.$$

Exercise 5.3. Consider the AR(2)

$$y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + \epsilon_t$$

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 tell you about y_t?

Consider the VAR(1)

$$\begin{aligned} x_t &= x_{t-1} + \epsilon_{1,t} \\ y_t &= \beta x_{t-1} + \epsilon_{2,t} \end{aligned}$$

where $\{\boldsymbol{\epsilon}_t\}$ 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)

$$\begin{bmatrix} x_t \\ y_t \end{bmatrix} = \begin{bmatrix} 0.4 & 0.3 \\ 0.8 & 0.6 \end{bmatrix} \begin{bmatrix} x_{t-1} \\ y_{t-1} \end{bmatrix} + \begin{bmatrix} \epsilon_{1,t} \\ \epsilon_{2,t} \end{bmatrix}$$

where $\{\boldsymbol{\epsilon}_t\}$ is a vector white noise process.

i. How can you verify that x_t and y_t are cointegrated?

ii. Write this model in error correction form.

iii. Compute the speed of adjustment coefficient $\boldsymbol{\alpha}$ and the cointegrating vector $\boldsymbol{\beta}$ where the $\boldsymbol{\beta}$ on x_t is normalized to 1.

Exercise 5.4. Data on interest rates on US government debt was collected for 3-month ($3MO$) T-bills, and 3-year ($3YR$) and 10-year ($10YR$) bonds from 1957 until 2009. Using these three series, the following variables were defined

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 was performed on this set and the p-vals were

	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 5.6 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 "0" lag impulses 0 while other aren't?

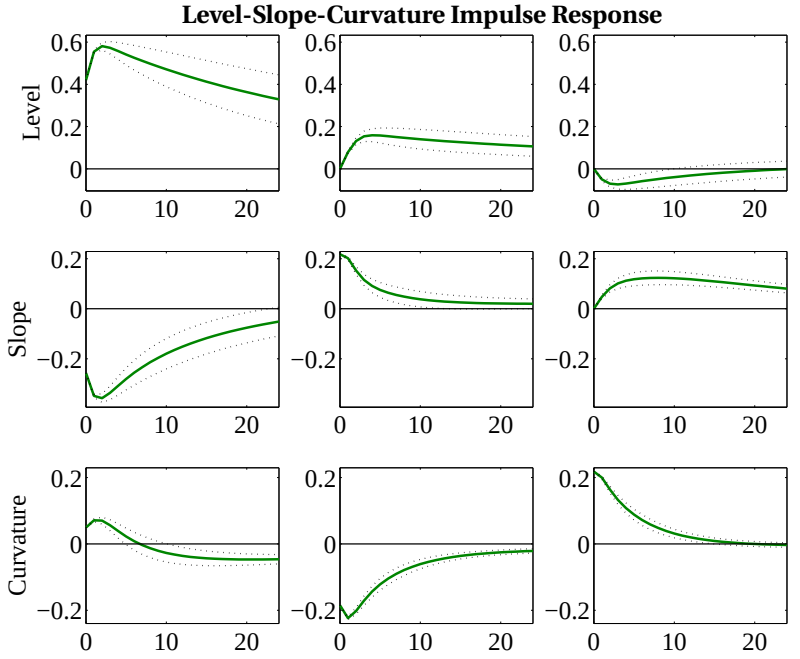

Figure 5.6: Impulse response functions and 95% confidence intervals for the level-slope-curvature exercise.

Exercise 5.5. Answer the following questions:

i. Consider the AR(2)

$$y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + \epsilon_t$$

(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 tell you about y_t?

ii. Consider the VAR(1)

$$\begin{aligned} x_t &= x_{t-1} + \epsilon_{1,t} \\ y_t &= \beta x_{t-1} + \epsilon_{2,t} \end{aligned}$$

where $\{\boldsymbol{\epsilon}_t\}$ 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)

$$\begin{bmatrix} x_t \\ y_t \end{bmatrix} = \begin{bmatrix} 0.625 & -0.3125 \\ -0.75 & 0.375 \end{bmatrix} \begin{bmatrix} x_{t-1} \\ y_{t-1} \end{bmatrix} + \begin{bmatrix} \epsilon_{1,t} \\ \epsilon_{2,t} \end{bmatrix}$$

where $\{\boldsymbol{\epsilon}_t\}$ is a vector white noise process.

(a) How can you verify that x_t and y_t are cointegrated?

(b) Write this model in error correction form.

(c) Compute the speed of adjustment coefficient $\boldsymbol{\alpha}$ and the cointegrating vector $\boldsymbol{\beta}$ where the $\boldsymbol{\beta}$ 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 $\epsilon_t = \rho\epsilon_{t-1} + \nu_t$ where $\{\nu_t\}$ 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.

$$\begin{aligned} x_t &= \eta_{1,t} + \theta_{11}\eta_{1,t-1} + \theta_{12}\eta_{2,t-1} \\ y_t &= \eta_{2,t} + \theta_{21}\eta_{1,t-1} + \theta_{22}\eta_{2,t-1} \end{aligned}$$

iv. Define the autocovariance matrix of a vector process.

v. Compute the autocovariance matrix $\boldsymbol{\Gamma}_j$ for $j = 0, \pm 1$.

vi. The AIC, HQC and SBIC were 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?

Lag Length	AIC	HQC	SBIC
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)

$$x_t = x_{t-1} + \epsilon_{1,t}$$
$$y_t = \beta x_{t-1} + \epsilon_{2,t}$$

where $\{\boldsymbol{\epsilon}_t\}$ 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)

$$\mathbf{y_t} = \mathbf{\Phi_1 y_{t-1}} + \mathbf{\Phi_2 y_{t-2}} + \boldsymbol{\epsilon_t}.$$

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)

$$\mathbf{y_t} = \mathbf{\Phi_1 y_{t-1}} + \boldsymbol{\epsilon_t}.$$

Define Granger Causality and explain what conditions on $\mathbf{\Phi}_1$ 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 $\mathbf{\Phi}_1$ 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 5.9. Consider a VAR(1)

$$\boldsymbol{y}_t = \mathbf{\Phi}_1 \boldsymbol{y}_{t-1} + \boldsymbol{\epsilon}_t$$

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

ii. Define cointegration for this model.

iii. What conditions on the eigenvalues of $\mathbf{\Phi}_1$ are required for cointegration to be present?

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

$$\Delta \boldsymbol{y}_t = \mathbf{\Pi} \boldsymbol{y}_{t-1} + \boldsymbol{\epsilon}_t.$$

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

(a) $\boldsymbol{y}_{t-1}$ are stationary

(b) $\boldsymbol{y}_{t-1}$ are cointegrated

(c) $\boldsymbol{y}_{t-1}$ are random walks

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