10th Polish-American Summer School in Economics
Forecasting The Polish Economy

Notes In Econometric Time-Series Estimation

Bernard J. Morzuch
Department of Resource Economics
University of Massachusetts
Amherst, Massachusetts

July 2-15, 2000

Stability and Stationarity

① Everything we've done so far had to do with the series becoming *stable* or *converging*.

②

↓

$$y_t = a y_{t-1} + b_t$$

$$y_1 = a y_0 + b_1$$
$$y_2 = a y_1 + b_2$$
$$\vdots \quad \vdots \quad \vdots$$
$$y_j = a y_{j-1} + b_j$$

Through repeated substitution:

$$y_t = a^t y_0 + \sum_{k=1}^{t} a^{t-k} b_k$$

↓

This is the *general solution.*

↓

③ The series becoming *stable* or *converging* depends upon the magnitude of "a" relative to the value 1

⇒

④

Suppose that b_t is replaced with ε_t, a stochastic disturbance (white noise), assumed to have nice properties:

- $E(\varepsilon_t) = E(\varepsilon_{t-1}) = \dots = 0$
- $E(\varepsilon_t^2) = E(\varepsilon_{t-1}^2) = \dots = \sigma^2$
- $E(\varepsilon_t \varepsilon_{t-s}) = E(\varepsilon_{t-j} \varepsilon_{t-j-s}) = 0$ for all s

Enders, Chapter 2, page 66

⇒

⑤

$$y_t = a y_{t-1} + \varepsilon_t$$

$$y_1 = a y_0 + \varepsilon_1$$
$$y_2 = a y_1 + \varepsilon_2$$
$$\vdots \quad \vdots \quad \vdots$$
$$y_j = a y_{j-1} + \varepsilon_j$$

Through repeated substitution:

$$y_t = a^t y_0 + \sum_{k=1}^{t} a^{t-k} \varepsilon_k$$

↓

This is the *general solution*.

↓

This is similar to Enders, page 70, equation 2.10

↓

⑥ Note: Introducing explicitly the stochastic disturbance ε_t puts y_t in the position of being a random variable. So, explore the behavior of y_t through expected value operations.

Additional, new concept.

y_t is *weakly stationary* or *covariance stationary* if

$$E(y_t) = E(y_{t-s}) = \mu \quad (<\infty)$$
$$E[(y_t - \mu)^2] = E[(y_{t-s} - \mu)^2] = \sigma_y^2$$
$$E[(y_t - \mu)(y_{t-s} - \mu)] = E[(y_{t-j} - \mu)(y_{t-j-s} - \mu)] = \gamma_s \text{ for all } s.$$

↓

Enders, page 69

⑦ ⇐ A problem ⇐

⑧ $E(y_t) \neq E(y_{t+s}) \rightarrow$ The y sequence cannot be *stationary* (Enders, page 70)

However, if t is large, $E(y_t)$ approaches $E(y_{t+s})$. Similar reasoning applies to the variance and autocovariance. (Enders, page 71)

⑨ "Thus, if a sample is generated by a process that has recently begun, the realizations may not be *stationary*.

⑩ Relationship: Enders, p. 68: The *stability* condition is a necessary condition for the time series to be *stationary*

How to put this together?

- You start with a series.
- You certainly don't begin by looking at any "coefficients". After all, it is these that you will eventually estimate.
- You do begin by testing for stationarity. We will use formal tests and several helpful graphical displays.
    - What happens if the series fails the stationarity test? Transformations exist to achieve stationarity.
- Once the series is stationary, proceed with estimation.

Other preliminary issue: simple equation $y_t = a_1 y_{t-1} + \varepsilon_t$

- More lags in y?
    - e.g., $y_{t-2},\ y_{t-3},\ \dots\ y_{t-p}$ and respective coefficients $a_2,\ a_3,\ \dots,\ a_p$
    - called autoregressive (AR) components
- Lags in ε?
    - e.g., $\varepsilon_{t-1},\ \varepsilon_{t-2},\ \dots\ \varepsilon_{t-q}$ and respective coefficients $\beta_1,\ \beta_2,\ \dots,\ \beta_q$
    - called moving average (MA) components
- New model may be:

$$y_t = a_1 y_{t-1} + a_2 y_{t-2} + \dots + a_p y_{t-p} + \varepsilon_t + \beta_1 \varepsilon_{t-1} + \beta_2 \varepsilon_{t-2} + \dots + \beta_q \varepsilon_{t-q}$$

which is an ARMA (p, q) model.

This equation appears at the top of page 63 in Enders.

Regression-Based Forecasting Models

Causal model:

$$Y_t = \beta_0 + \beta_1 X_t + \varepsilon_t$$

Prediction in t+1: $\hat{Y}_{t+1} = b_0 + b_1 X_{t+1}$

X must be predicted in t+1.

Trend-based forecasting model:

$Y_t = \beta_0 + \beta_1 t + \varepsilon_t$, where t is time.

Prediction in t+1: $\hat{Y}_{t+1} = b_0 + b_1(t+1)$

t does not have to be predicted in t+1.

Regression model with an indicator variable:

$$Y_t = \beta_0 + \beta_1 X_{t-1} + \varepsilon_t$$

Indicator variable X precedes Y in time (X_{t-1} comes before Y_t and is referred to as a leading indicator).

Prediction in t+1: $\hat{Y}_{t+1} = b_0 + b_1 X_t$

X_t does not have to be predicted.

Autoregressive model:

$$Y_t = \beta_0 + \beta_1 Y_{t-1} + \varepsilon_t$$

Prediction in t+1: $\hat{Y}_{t+1} = b_0 + b_1 Y_t$

Y_t does not have to be predicted.

Combined autoregressive, trend-projection, and indicator variable regression model:

$$Y_t = \beta_0 + \beta_1 t + \beta_2 X_{t-1} + \beta_3 Y_{t-1} + \varepsilon_t$$

Prediction in t+1: $\hat{Y}_{t+1} = b_0 + b_1(t+1) + b_2 X_t + b_3 Y_t$

The Unit Root Problem

- Revisit the *level* specification of the model:

$$Y_t = \beta_1 + \beta_2 X_t + \varepsilon_t$$

- In the case of first-order autoregressive errors, we have $\varepsilon_t = \rho\varepsilon_{t-1} + \mu_t$ where ρ is between -1 and +1.

- If ρ is outside of these limits, the error behaves in an explosive fashion.

- Q: What happens if $\rho = 1$?

- A: The error term becomes $\varepsilon_t = \varepsilon_{t-1} + \mu_t$. This *is* the random walk model.

  "It can be shown that the variance of ε_t goes to infinity as t goes to infinity."

  This *situation* is called the unit root problem.

Q: How to deal with this situation?

A: General rho-transformation to correct for autocorrelation

$$Y_t - \rho Y_{t-1} = \beta_1(1-\rho) + \beta_2(X_t - \rho X_{t-1}) + \mu_t .$$

With $\rho = 1$, the above becomes

$$Y_t - Y_{t-1} = \beta_2(X_t - X_{t-1}) + \mu_t \text{ where } \mu_t \text{ is well behaved.}$$

- The purpose in doing this is to illustrate how the problem can be solved using simple differencing. Notice that ordinary least squares regression (OLS) can be used on this last equation when it was not proper to use it on its *level* counterpart.

- In the case of general-order autocorrelation, the unit root problem occurs when the sum of the ρ's is equal to 1.

Spurious Regressions: The Background For Unit Root Tests

- Univariate time series requirement: stationarity.
- Differencing: method of achieving stationarity.
- Same requirement holds for variables in a vector autoregression (VAR) model *or* in any single equation, regression-based model.
- What happens if variables in such models are nonstationary?
- Regression results may be "spurious".
    - They look good: High R^2 ; significant t-values.
    - But, they have no real meaning.
- Characteristics of such a model:
    - $Y_t = \beta_1 + \beta_2 X_t + \varepsilon_t$ $\rightarrow$ contemporaneous relationships.
    - Y_t and X_t are trended or nonstationary.
    - X_t is a random regressor (e.g., its values are simultaneously determined with Y_t).
    - These last two points -- random regressor and nonstationarity -- result in least squares estimators not being consistent.
- Granger and Newbold illustrate the problem with a Monte Carlo experiment.
    - Start: Y_t, X_t generated as *independent random walks*.
    - Recall Z_t is a random walk if:

$$Z_t = Z_{t-1} + \varepsilon_t \text{ where } \varepsilon_t \sim (0, \sigma^2)$$

    - Recall also that a random walk is a relatively smooth series that changes slowly. Many economic series have random walk characteristics.
- They created repeated, independent random samples of size t = 50 observations on Y_t and X_t.

- They ran the model: $Y_t = \beta_1 + \beta_2 X_t + \varepsilon_t$
- H_0: $\beta_2 = 0$; H_A: $\beta_2 \neq 0$
- Q: Which hypothesis do you think would tend to be supported the majority of the time?
- A: H_0: $\beta_2 = 0$
- What they found: H_0: $\beta_2 = 0$ was rejected 75% of the time at a level $\alpha = 0.05$!

  ⇒ A significant relationship was found where none existed in 3/4 of the cases.

- Other conclusions:

  When regression equations like $Y_t = \beta_1 + \beta_2 X_t + \varepsilon_t$ are specified between *levels* of economic time series that behave like random walks, these equations frequently have *high* R^2 values and *low* Durbin-Watson statistics.

  ⇒ usual t tests are misleading; they reject H_0 of no relationship too often.

  ⇒ usual t tests accept as significant relationships that are spurious far too often.

- So, when a regression leads to *high* R^2 and *low* Durbin-Watson statistic, the relationship should be estimated in *first differences* of the variables rather than *levels*:

$$\Delta Y_t = \gamma_1 + \gamma_2 \Delta X_t + \upsilon_t$$

  Monte Carlo results of this estimation do not lead to overacceptance of spurious relationships as significant ones.

$y_t = a_0 + a_1 z_t + e_t$ (4.5) $y_t = y_{t-1} + \epsilon_{yt}$ (4.6) $z_t = z_{t-1} + \epsilon_{zt}$ (4.7)

In their Monte Carlo analysis, Granger and Newbold generated many such samples and for each sample estimated a regression in the form of (4.5). Since the $\{y_t\}$ and $\{z_t\}$ sequences are independent of each other, Equation (4.5) is necessarily meaningless; any relationship between the two variables is spurious. Surprisingly, at the 5% significance level, they were able to reject the null hypothesis $a_1 = 0$ in approximately 75% of the time. Moreover, the regressions usually had very high R^2 values and the estimated residuals exhibited a high degree of autocorrelation.

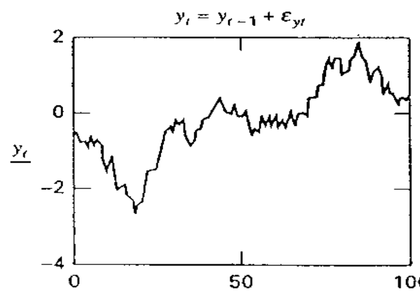
The simplest way to examine the properties of the $\{e_t\}$ sequence is to abstract from the intercept term a_0 and rewrite (4.5) as

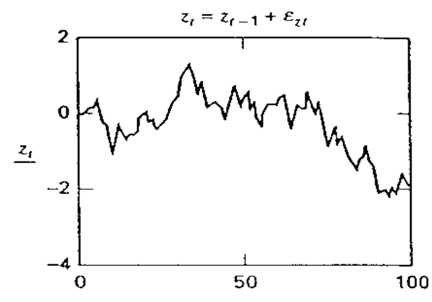
$$e_t = y_t - a_1 z_t$$

If z_t and y_t are generated by (4.6) and (4.7), we can impose the initial conditions $y_0 = x_0 = 0$, so that

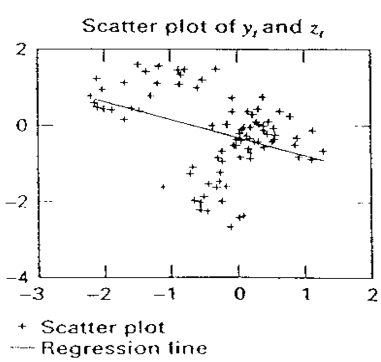
$$e_t = \sum_{i=1}^{t} \epsilon_{yi} - a_1 \sum_{i=1}^{t} \epsilon_{zi} \tag{4.8}$$

Clearly, the variance of the error becomes infinitely large as t increases.





Since the $\{\epsilon_{yt}\}$ and $\{\epsilon_{zt}\}$ sequences are independent, the regression of y_t on z_t is spurious. Given the realizations of the random disturbances, it appears as if the two sequences are related. In the scatter plot of y_t against z_t, you can see that y_t tends to rise as z_t decreases. The regression equation of y_t on z_t will capture this tendency. The correlation coefficient between y_t and z_t is -0.372 and a linear regression yields $y_t = -0.46z_t - 0.31$. However, the residuals from the regression equation are nonstationary.



The Unit Root Problem: Setting Up The Appropriate Hypotheses And Using The Appropriate Test Statistic

Suppose that the data generating process (DGP) is:

$$y_t = 1y_{t-1} + \varepsilon_t. \qquad (1)$$

but all we have to go on is

$$y_t = a_1 y_{t-1} + \varepsilon_t. \qquad (2)$$

The traditional approach to hypothesis testing is to test the null that a_1 equals zero versus the alternative that a_1 does not equal zero. If, in fact, $a_1 = 0$ under the null, this implies that ε_t in equation (2) is well-behaved under this same hypothesis. Specifically, var ε_t is finite; it is assumed not to explode as t increases. Given the characteristics of the model under this hypothesis, the test statistic chosen to evaluate the null hypothesis must abide by these characteristics. Such a test statistic is t.

Q: If H_0 is rejected using this traditional approach, does this provide us with evidence that $a_1 = 1$; i.e., that we have a unit root process?
A: No! We merely have evidence to suggest that $a_1 \neq 0$.

If our motive for hypothesis testing is to test for the presence of a unit root, i.e., that $a_1 = 1$, we should set up our null and alternative to reflect this objective. Thus the null now becomes a_1 equals one versus the alternative that a_1 does not equal one. If, in fact, $a_1 = 1$ under the null, this implies that ε_t in equation (1) is not well-behaved under this same hypothesis. Specifically, var ε_t goes to infinity as t increases when equation (1) is the appropriate model. This characteristic makes inappropriate the choice of the t-test as the mechanism for evaluating the null hypothesis.

If we knew the theoretical distribution of a_1 under the null hypothesis of a unit root, we could perform a test of significance.

Q: How do we formally test for the presence of a unit root?
A: Dickey and Fuller developed the appropriate test statistic.

A Closer Look At The Types Of Hypothesis Tests On Coefficients In An Autoregressive Model

Suppose we know that a series is generated by:

$$y_t = a_1 y_{t-1} + \varepsilon_t$$

where ε_t is white noise. (This is a typical set-up for an autoregressive process).

Test 1: Test the null hypothesis: H_0: $a_1=0$

Under H_0 true, $|a_1|<1$. This, along with ε_t being white noise, guarantee that y_t is stationary. Under H_0 true, OLS is valid; this problem has all of the characteristics that make a t-test valid.

Test 2: Test the null hypothesis: H_0: $a_1=1$

Under H_0 true, y_t is nonstationary. (It is a random walk process). The variance of y_t becomes infinitely large as t increases. Under H_0 true, it is inappropriate to use classical statistical methods to estimate and perform significance tests on a_1.

Sources On The Formal Procedure To Test For The Presence Of A Unit Root

Dickey, David and Wayne A. Fuller. "Distribution of the Estimates for Autoregressive Time Series with a Unit Root." *Journal of the American Statistical Association*. 74(June 1979), 427-31.

Dickey, David and Wayne A. Fuller. "Likelihood Ratio Statistics for Autoregressive Time Series with a Unit Root." *Econometrica*. 49(July 1981), 1057-72.

Procedure To Test For Unit Roots

• Competing Models:	Pure Random Walk (Unit Root Process)	vs	Stationary AR(1) Process (No Unit Root)

- Model Form: $y_t = a_1 y_{t-1} + \varepsilon_t$

- Appropriate Hypotheses *When Testing For A Unit Root*:

  $H_0\!: a_1 = 1$

  $H_A\!: a_1 < 1$

- Actual Estimating Equation Used To Conduct The Hypothesis Test: UROOT command in TSP (pages 16-1 to 16-5)

  START: $y_t = a_1 y_{t-1} + \varepsilon_t$

  Subtract y_{t-1} From Both Sides: $y_t - y_{t-1} = a_1 y_{t-1} - y_{t-1} + \varepsilon_t$

  Collect Terms: $\Delta y_t = (a_1 - 1) y_{t-1} + \varepsilon_t$

  $\Delta y_t = \gamma \; y_{t-1} + \varepsilon_t$

  So, $H_0\!: a_1 = 1$ is equivalent to $H_0\!: \gamma = 0$

- Dickey and Fuller developed the *test statistic* to reflect the conditions of the null hypothesis being examined: τ

Models Of Increasing Complexity

Form	Test Statistic
• Start: $\Delta y_t = \gamma y_{t-1} + \varepsilon_t$	τ
• Random walk with drift (include an intercept) $\Delta y_t = a_0 + \gamma y_{t-1} + \varepsilon_t$	τ_μ
• Random walk with drift and linear time trend $\Delta y_t = a_0 + \gamma y_{t-1} + a_2 t + \varepsilon_t$	τ_τ
• Augmented Dickey-Fuller model (include stationary forms of lagged variables)	

- $\Delta y_t = a_0 + \gamma y_{t-1} + a_2 t + \beta_1 \Delta y_{t-1} + \beta_2 \Delta y_{t-2} + \varepsilon_t$

Possible Forecasting Approaches

<u>Time-Series Approaches</u>

Univariate: Exponential Smoothing

MA

AR

ARMA

Multivariate: Transfer Functions

<u>Regression-Based Approaches</u>

Single Equation Estimation

Systems of Equations:

- Simultaneous Systems
  - Deriving the Reduced Form
  - Using the Reduced Form For Forecasting
- Vector Autoregression

INTRODUCTION TO VAR ANALYSIS

Consider, the simple bivariate system:

$$y_t = b_{10} - b_{12}z_t + \gamma_{11}y_{t-1} + \gamma_{12}z_{t-1} + \epsilon_{yt} \quad (5.19)$$
$$z_t = b_{20} - b_{21}y_t + \gamma_{21}y_{t-1} + \gamma_{22}z_{t-1} + \epsilon_{zt} \quad (5.20)$$

Equations (5.19) and (5.20) constitute a *first-order* vector autoregression (VAR) since the longest lag length is unity. This simple two-variable first-order VAR is useful for illustrating the multivariate higher-order systems that are introduced in Section 8. The structure of the system incorporates feedback since y_t and z_t are allowed to affect each other.

Equations (5.19) and (5.20) are not reduced-form equations since y_t has a contemporaneous effect on z_t and z_t has a contemporaneous effect on y_t. Fortunately, it is possible to transform the system of equations into a more usable form. Using matrix algebra, we can write the system in the compact form:

$$\begin{bmatrix} 1 & b_{12} \\ b_{21} & 1 \end{bmatrix} \begin{bmatrix} y_t \\ z_t \end{bmatrix} = \begin{bmatrix} b_{10} \\ b_{20} \end{bmatrix} + \begin{bmatrix} \gamma_{11} & \gamma_{12} \\ \gamma_{21} & \gamma_{22} \end{bmatrix} \begin{bmatrix} y_{t-1} \\ z_{t-1} \end{bmatrix} + \begin{bmatrix} \epsilon_{yt} \\ \epsilon_{zt} \end{bmatrix}$$

or

$$Bx_t = \Gamma_0 + \Gamma_1 x_{t-1} + \epsilon_t$$

Premultiplication by B^{-1} allows us to obtain the vector autoregressive (VAR) model in *standard* form:

$$x_t = A_0 + A_1 x_{t-1} + e_t \quad (5.21)$$

Using this new notation, we can rewrite (5.21) in the equivalent form:

$$y_t = a_{10} + a_{11}y_{t-1} + a_{12}z_{t-1} + e_{1t} \quad (5.22a)$$
$$z_t = a_{20} + a_{21}y_{t-1} + a_{22}z_{t-1} + e_{2t} \quad (5.22b)$$

To distinguish between the systems represented by (5.19) and (5.20) versus (5.22a) and (5.22b), the first is called a structural VAR or the primitive system and the second is called a VAR in standard form. It is important to note that the error terms (i.e., e_{1t} and e_{2t}) are composites of the two shocks ϵ_{yt} and ϵ_{zt}. Since $e_t = B^{-1}\epsilon_t$, we can compute e_{1t} and e_{2t} as

$$e_{1t} = (\epsilon_{yt} - b_{12}\epsilon_{zt})/(1 - b_{12}b_{21}) \quad (5.23)$$

$$e_{2t} = (\epsilon_{zt} - b_{21}\epsilon_{yt})/(1 - b_{12}b_{21}) \quad (5.24)$$

Since ϵ_{yt} and ϵ_{zt} are white-noise processes, it follows that both e_{1t} and e_{2t} have zero means, constant variances, and are individually serially uncorrelated.

Mechanics of Vector Autoregression

Recall the autoregressive model for Y whose value Y_t depends on past values up to lag length P.

- *Assume* the stochastic process to be stationary. (If it weren't, some correction is done).
- This is an AR(P) process.
- Form:

$$Y_t = \varphi_1 Y_{t-1} + \varphi_2 Y_{t-2} + \dots + \varphi_p Y_{t-p} + \varepsilon_t$$

Recall:

ε_t : white noise, i.e.,

$$E(\varepsilon_t) = 0$$
$$E(\varepsilon_t \varepsilon_s) = 0 \quad \text{for} \quad s \neq t \rightarrow \text{zero autocovariance}$$
$$E(\varepsilon_t \varepsilon_t) = \sigma_\varepsilon^2 \rightarrow \text{constant variance}$$

Q: *Suppose* we *know* the order P (we have yet to find it). How can we get estimates of $\varphi_1, \varphi_2, \dots, \varphi_p$?

A: Rather than use the nonlinear techniques previously suggested for time series models, why not use ordinary least squares? Notice that this equation is in the form of the linear statistical model.

$$Y_{p+1} = \varphi_1 Y_p + \dots + \varphi_p Y_1 + \varepsilon_{p+1}$$
$$Y_{p+2} = \varphi_1 Y_{p+1} + \dots + \varphi_p Y_2 + \varepsilon_{p+2}$$
$$\vdots$$
$$Y_t = \varphi_1 Y_{t-1} + \dots + \varphi_p Y_{t-p} + \varepsilon_t$$

Note:
- Each equation has "P" φ's to be estimated. (In an actual problem, we have yet to determine P).
- The subscript of the dependent variable matches its error.
- The number of right-hand-side Y's is P.
- The (1,1) element of the dependent variable vector is 1 more than P.

In matrix notation:

$$\begin{bmatrix} Y_{p+1} \\ Y_{p+2} \\ \vdots \\ Y_t \end{bmatrix} = \begin{bmatrix} Y_p & Y_{p-1} & \cdots & Y_1 \\ Y_{p+1} & Y_p & \cdots & Y_2 \\ \vdots & \vdots & \ddots & \vdots \\ Y_{t-1} & Y_{t-2} & \cdots & Y_{t-p} \end{bmatrix} \begin{bmatrix} \varphi_1 \\ \varphi_2 \\ \vdots \\ \varphi_p \end{bmatrix} + \begin{bmatrix} \varepsilon_{p+1} \\ \varepsilon_{p+2} \\ \vdots \\ \varepsilon_t \end{bmatrix}$$

$$\downarrow \qquad\qquad \downarrow \qquad\qquad \downarrow \qquad\qquad \downarrow$$

$$\underline{Y}_p = Y_p \; \underline{\varphi}_p + \underline{\varepsilon}$$

Estimating equations are:

- $\hat{\underline{\varphi}}_p = (Y_p' Y_p)^{-1} \; Y_p' \underline{Y}_p$
- $\Sigma_{\hat{\underline{\varphi}}\hat{\underline{\varphi}}} = \hat{\sigma}_e^2 \; (Y_p' Y_p)^{-1}$

Where:

$$\hat{\sigma}_e^2 = \frac{(\underline{Y}_p - Y_p \hat{\underline{\varphi}}_p)' \; (\underline{Y}_p - Y_p \hat{\varphi}_p)}{T - 2_p}$$

Q: Why 2p in denominator rather than P?

A: Because we treat $Y_1, ..., Y_p$ as presample values *and* estimate P parameters.

All of the above suggests a name for the procedure of estimating the parameters of this *autoregressive* model. In particular, with the data arranged in the familiar *vector* format of the classical linear *regression* model, we can refer to the above appropriately as vector autoregression.

Unfortunately, the order P of the AR process is not known in advance. How do we choose P?

One way to identify the order of an adequate AR process:

- Estimate processes of increasing order K
- Test the significance of φ_k → this coefficient is called the kth partial autocorrelation coefficient.
  - It measures the correlation between Y_t and Y_{t-k} not accounted for by an AR(K-1).
  - The sequence of partial autocorrelations is called the partial autocorrelation function.
- To test the significance of φ_k, we need to know the distribution of its estimate $\hat{\varphi}_k$, obtained as the last coordinate of

$$\hat{\underline{\varphi}}_k = (Y_k' Y_k)^{-1} \; Y_k' \underline{Y}_k$$

For large sample size, the estimated partial autocorrelations are approximately normally distributed with mean zero and variance 1/T, where T is sample size.

To check the significance of φ_k, form (for example) the 95% confidence intervals:

$$\left(\hat{\varphi}_k - \frac{2}{\sqrt{T}} \; , \; \hat{\varphi}_k + \frac{2}{\sqrt{T}} \right)$$

Equivalently, check whether the estimated partial autocorrelation falls within the two standard error bounds.

Suppose we take the previous univariate series and augment its matrix specification (vertically and horizontally) by another univatiate series (call this new series X_t).

Motivation: Think of

Y_t : GDP

X_t : Investment

Q: Suppose we want to forecast GDP. Given *just* what was presented on the previous pages, how would we proceed?

A: By using the information contained in past values of GDP *only*.

Q: Might not forecasts of GDP be improved if past values of investment are included in the model specification? Recall that GDP = f(investment).

A: Yes $\therefore$ augment specification horizontally.

Q: We know that investment and GDP may influence each other, so why not also use GDP to predict investment:

A: This sounds good, so let's augment specification vertically.

Q: Why do all this?

A: Admitting this information in this estimation procedure might improve the forecasts. The parallel in econometrics is the movement from a single equation model to a simultaneous equations model.

An important issue surfaces as a result of moving in this direction ... that of *causality*.

Q: Does GDP cause investment, or investment cause GDP, or is it some combination of the two?

A: Difficult issue; this will be addressed later.

For now, let's look at the augmented specification without considering causality.

Let there be a stochastic process (assumed to be autoregressive) for Y_t and X_t.

Y_t has been presented and its specification completely laid out. Let's do the same thing for X_t.

$$X_t = \dot{\varphi}_1 X_{t-1} + \dot{\varphi}_2 X_{t-2} + \ldots + \dot{\varphi}_1 X_{t-p} + \omega_t \Rightarrow \text{ an AR(P) process}$$

Also,

ω_t : white noise

$E(\omega_t) = 0$

$E(\omega_t \omega_s) = 0 \quad \text{for} \quad s \neq t$

$E(\omega_t \omega_t) = \sigma_\omega^2$

Let there be this "back and forth" influence. Now, develop the vertically and horizontally augmented system.

Introduce the following notation on the φ:

$$\varphi_{ij,k}$$

where:

i = Equation number (= 1 or 2)

Equation 1 if left-hand-side variable is Y_t
Equation 2 if left-hand-side variable is X_t

j = Type of right-hand-side variables included (= 1 or 2)

1 if corresponding right-hand-side variable is a Y_t
2 if corresponding right-hand-side variable is a X_t

k = Order of the process for the respective right-hand-side variable

$k = 1, \ldots, p$ for Y_t
$k = 1, \ldots, m$ for X_t (For now, assume that p=m)

System is:

$$Y_t = \varphi_{11,1} Y_{t-1} + \varphi_{11,2} Y_{t-2} + \dots + \varphi_{11,p} Y_{t-p} + \varphi_{12,1} X_{t-1} + \varphi_{12,2} X_{t-2} + \dots + \varphi_{12,m} X_{t-m} + \varepsilon_t$$

$$X_t = \varphi_{21,1} Y_{t-1} + \varphi_{21,2} Y_{t-2} + \dots + \varphi_{21,p} Y_{t-p} + \varphi_{22,1} X_{t-1} + \varphi_{22,2} X_{t-2} + \dots + \varphi_{22,m} X_{t-m} + \omega_t$$

In matrix notation:

$$\begin{matrix} Y_t \\ \\ \\ X_t \end{matrix} \quad \begin{bmatrix} Y_{p+1} \\ Y_{p+2} \\ \vdots \\ Y_t \\ \bullet\bullet\bullet \\ X_{m+1} \\ X_{m+2} \\ \vdots \\ X_t \end{bmatrix} = \begin{bmatrix} \begin{matrix} Y_p & Y_{p-1} & \cdots & Y_1 & X_p & X_{p-1} & \cdots & X_1 \\ Y_{p+1} & Y_p & \cdots & Y_2 & X_{p+1} & X_p & \cdots & X_2 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ Y_{t-1} & Y_{t-2} & \cdots & Y_{t-p} & X_{t-1} & X_{t-2} & \cdots & X_{t-p} \end{matrix} & 0 \\ 0 & \begin{matrix} Y_p & Y_{p-1} & \cdots & Y_1 & X_p & X_{p-1} & \cdots & X_1 \\ Y_{p+1} & Y_p & \cdots & Y_2 & X_{p+1} & X_p & \cdots & X_2 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ Y_{t-1} & Y_{t-2} & \cdots & Y_{t-p} & X_{t-1} & X_{t-2} & \cdots & X_{t-p} \end{matrix} \end{bmatrix} \begin{bmatrix} \varphi_{11,1} \\ \varphi_{11,2} \\ \vdots \\ \varphi_{11,p} \\ \varphi_{12,1} \\ \varphi_{12,2} \\ \vdots \\ \varphi_{12,p} \\ \bullet\bullet\bullet \\ \varphi_{21,1} \\ \varphi_{21,2} \\ \vdots \\ \varphi_{21,p} \\ \varphi_{22,1} \\ \varphi_{22,2} \\ \vdots \\ \varphi_{22,p} \end{bmatrix} + \begin{bmatrix} \varepsilon_{p+1} \\ \varepsilon_{p+2} \\ \vdots \\ \varepsilon_t \\ \bullet\bullet\bullet \\ \omega_{p+1} \\ \omega_{p+2} \\ \vdots \\ \omega_t \end{bmatrix}$$

$$\underline{Y} = X \quad \underline{\beta} + \underline{\mu}$$

Characteristics:

- Notice that the previous model for Y_t is merely augmented by X_t. (Notice, we are assuming that Y_t is AR(P) and X_t AR(M) = AR(P); we have yet to discover the process of each).

- The grand X matrix is block diagonal. If M = P (as is presently assumed), the two main diagonal blocks are precisely the same. (This has implications for the estimation procedure to use).

- This *is* a special case of a simultaneous system.

  - Since all the R.H.S. values of X are lagged, each column is predetermined.

  - The structural form and reduced form coincide.

  - Each equation (Y_t and X_t) can be estimated by OLS.

- Notice: ε_{p+1} may be correlated with ω_{p+1}, ε_{p+2} with ω_{p+2} ... ε_t with ω_t.

- Q: Why not use an estimation technique that makes use of the fact that the errors are mutually correlated, e.g., seemingly unrelated regression (SUR) [also known as Zellner estimation]? This could improve the efficiency of the estimators.

  A: Efficiency improves only if errors are in fact correlated and the respective columns of X are uncorrelated. (This is not the case here.)

Furthermore

It is not always useful to estimate the autoregressive model in the previous form

$$\underline{Y} = X\,\underline{\beta} + \underline{\mu}\ .$$

Q: Why?

A: Frequently some of the $\varphi_{ij,k}$'s will be zero. Neglecting valid zero restrictions results in inefficient estimates.

Notice, for example, that the stochastic process for Y_t may very well have different lag structures for *both* Y_{t-1} and X_{t-1} *than* X_t has for Y_{t-1} and X_{t-1}. This means that the 2 block diagonal matrices in X are now different. Taking account of the correlation in the error terms now improves the efficiency of the estimators.

The variance-covariance matrix of the error term is:

$$E\begin{bmatrix}\varepsilon_t\\ \omega_t\end{bmatrix}\begin{bmatrix}\varepsilon_t, & \omega_t\end{bmatrix} = \begin{bmatrix}\sigma_\varepsilon^2 & \varepsilon_{\varepsilon\omega}\\ \sigma_{\varepsilon\omega} & \sigma_\omega^2\end{bmatrix} = \Sigma_{\varepsilon\omega}$$

and the estimator of $\underline{\beta}$ is

$$\hat{\underline{\beta}} = (X'\Sigma_{\varepsilon\omega}^{-1}X)^{-1}X'\Sigma_{\varepsilon\omega}\underline{Y}$$

Q: How does one determine an adequate set of regressors for each equation?

A: In the univariate case, we based the choice on an adequate order of an AR process on the partial autocorrelations. For multivariate processes the partial autocorrelations are matrices.

Hsiao (JASA 74:553-560) suggests using Akaike's final prediction error (FPE) criterion; i.e., search for the combination "A" and "B" in each equation for which

$$FPE_{(A,B)} = \left(\frac{T + \text{no. of parameters}}{T - \text{no. of parameters}}\right)\left(\frac{SSE_{(A,B)}}{T}\right)$$

is a minimum.

Notation is as follows:

T: sample size used in the estimation.

$SSE_{(A,B)}$: sum of squared errors if $Y_{t-1}, ..., Y_{t-a}, X_{t-1}, ..., X_{t-b}$ are used as regressors in a particular equation.

Notice the effect of increasing the number of parameters on the right-hand-side of FPE.

- It increases the first term.
- It decreases the second term.

It is assumed that these two forces are balanced when their product reaches its minimum.

Actual implementation of FPE criterion:

- Specify a maximum number (MAX) for A & B that one is willing to consider.
- A univariate AR model is identified for Y_t by varying a from zero to MAX. (Suppose that $FPE_{(a,-)}$ obtains its minimum at $a = a_0$).
- Choose b between zero and MAX such that $b = b_0$ is the optimum. (Because a_0 was determined neglecting the x variable in the considered equation, it is possible that a_0 is unduly high as a result of an omitted variables effect). Thus, we compute $FPE_{(a,b_0)}$ for all $a = 0, 1, ..., a_0$ and use the minimizing combination (a_1, b_0).

Other methods may also identify adequate numbers of lags of X and Y. Too little is known about the performance of the various methods to attempt a ranking. (Hsiao's method is introduced here because of its simplicity).

Forecasting

Consider the previous model in matrix notation. Rearrange the model so that the parameters appear in the following block form:

$$\begin{bmatrix} Y_t \\ X_t \end{bmatrix} = \underbrace{\begin{bmatrix} \varphi_{11,1} & \varphi_{12,1} \\ \varphi_{21,1} & \varphi_{22,1} \end{bmatrix}}_{\Phi_1} \begin{bmatrix} Y_{t-1} \\ X_{t-1} \end{bmatrix} + \ldots + \underbrace{\begin{bmatrix} \varphi_{11,p} & \varphi_{12,p} \\ \varphi_{21,p} & \varphi_{22,p} \end{bmatrix}}_{\Phi_p} \begin{bmatrix} Y_{t-p} \\ X_{t-p} \end{bmatrix} + \begin{bmatrix} \varepsilon_t \\ \omega_t \end{bmatrix}$$

Now, let the one-step ahead forecast at time T be $\begin{bmatrix} Y_{T+1} \\ X_{T+1} \end{bmatrix}$

The computation for the one-step ahead forecast is:

$$\begin{bmatrix} Y_{T+1} \\ X_{T+1} \end{bmatrix} = \Phi_1 \begin{bmatrix} Y_T \\ X_T \end{bmatrix} + \ldots + \Phi_p \begin{bmatrix} Y_{T+1-p} \\ X_{T+1-p} \end{bmatrix}$$

The two-steps ahead forecast is:

$$\begin{bmatrix} Y_{T+2} \\ X_{T+2} \end{bmatrix} = \Phi_1 \begin{bmatrix} Y_{T+1} \\ X_{T+1} \end{bmatrix} + \ldots + \Phi_p \begin{bmatrix} Y_{T+2-p} \\ X_{T+2-p} \end{bmatrix}$$

where Y_{T+1} and X_{T+1} are the predictions from one-step ahead.

Prediction intervals for the forecasts can also be set up.

Granger Causality

Consider the model:

$$GDP = \beta_0 + \beta_1 \; GDP(-1) + \beta_2 \; GDP(-2) + \beta_3 \; M1(-1) + \beta_4 \; M1(-2) + \varepsilon$$

$$M1 = \alpha_0 + \alpha_1 \; GDP(-1) + \alpha_2 \; GDP(-2) + \alpha_3 \; M1(-1) + \alpha_4 \; M1(-2) + \mu$$

Definition: GDP is said to be Granger-caused by M1 if current and past information on M1 will *help* improve forecasts of GDP.

How to test? Look at first equation.

Test H_0: $\beta_3 = \beta_4 = 0$

- For the second equation, the task is to determine if M1 is Granger-caused by GDP.
- For this (second) equation, test H_0: $\alpha_1 = \alpha_2 = 0$

Note: Granger's concept of causality

- Does *not* imply a cause-effect relationship
- Is based on predictability

Results can be very sensitive to the number of lags included.

- H_0 may not be rejected for a small number of lags but is rejected for a large number of lags.

Likewise:

- H_0 may be rejected for a small number of lags but is not rejected for a large number of lags.

In practical investigations: More confidence in conclusions which are robust over different lag structures.

ESTIMATION AND IDENTIFICATION

One explicit aim of the Box–Jenkins approach is to provide a methodology that leads to parsimonious models. The ultimate objective of making accurate short-term forecasts is best served by purging insignificant parameter estimates from the model. Sims' (1980) criticisms of the "incredible identification restrictions" inherent in structural models argue for an alternative estimation strategy. Consider the following multivariate generalization of (5.21):

$$x_t = A_0 + A_1 x_{t-1} + A_2 x_{t-2} + \cdots + A_p x_{t-p} + e_t \tag{5.30}$$

where x_t = an $(n \times 1)$ vector containing each of the n variables included in the VAR

A_0 = an $(n \times 1)$ vector of intercept terms

A_i = $(n \times n)$ matrices of coefficients

and e_t = an $(n \times 1)$ vector of error terms

Sims' methodology entails little more than a determination of the appropriate variables to include in the VAR and a determination of the appropriate lag length. The variables to be included in the VAR are selected according to the relevant economic model. Lag-length tests (to be discussed below) select the appropriate lag length. Otherwise, no explicit attempt is made to "pare down" the number of parameter estimates. The matrix A_0 contains n intercept terms and each matrix A_i contains n^2 coefficients; hence, $n + pn^2$ terms need to be estimated. Unquestionably, a VAR will be *overparameterized* in that many of these coefficient estimates can be properly excluded from the model. However, the goal is to find the important interrelationships among the variables and not make short-term forecasts. Improperly imposing zero restrictions may waste important information. Moreover, the regressors are likely to be highly colinear, so that the t-tests on individual coefficients may not be reliable guides for paring down the model.

Note that the right-hand side of (5.30) contains only predetermined variables and the error terms are assumed to be serially uncorrelated with constant variance.[3] Hence, *each equation in the system can be estimated using OLS*. Moreover, OLS estimates are consistent and asymptotically efficient. Even though the errors are correlated across equations, seemingly unrelated regressions (SUR) do not add to the efficiency of the estimation procedure since both regressions have identical right-hand-side variables.

The issue of whether the variables in a VAR need to be stationary exists. Sims (1980) and others, such as Doan (1992), recommend against differencing *even if the variables contain a unit root*. They argue that the goal of VAR analysis is to determine the interrelationships among the variables, *not* the parameter estimates. The main argument against differencing is that it "throws away" information concerning the comovements in the data (such as the possibility of cointegrating relationships). Similarly, it is argued that the data need not be detrended. In a VAR, a trending variable will be well approximated by a unit root plus drift. However, the majority view is that the form of the variables in the VAR should mimic the true data-generating process. This is particularly true if the aim is to estimate a structural model. We return to these issues in the next chapter; for now, it is assumed that all variables are stationary. Two sets of questions at the end of this chapter ask you to compare a VAR in levels to a VAR in first differences.

Hsiao's paper: Start with definitions:

- M1: money supply = currency, traveler's checks, demand deposits
- M2: M1 + small denomination time and savings deposits + money market accounts
- GNP: output (or income)

Rationale for two-way causality:

- Monetarists say money is an independent driving force so that GNP = f(M1).
- Monetarist critics say money is a passive adapter to business conditions so that M1 = f(GNP).

P. 554, equation (2.5):

$$\hat{Y}_t = \hat{\Psi}^m_{11}(L)y_t + \hat{\Psi}^n_{12}(L)x_t + \hat{a}$$

Note: There is a similar equation with $\hat{X}_t$ appearing on the left-hand side.

m: order of lags in y_t
n: order of lags in x_t
â: estimated intercept

m+n+1 = number of estimated parameters.

Q: How should we evaluate this model relative to other models with different orders of lags?
A: Use equation (2.6): start here

$$FPE_y(m,n) = \frac{T+m+n+1}{T-m-n-1} \cdot \sum_{t=1}^{T} (Y_t - \hat{Y}_t)^2/T$$

An adjustment factor: It gets larger as more parameters are included.	Gets smaller as more parameters are included (Residual sum of squares becomes smaller as the model becomes more unrestricted.)
↑	↓
For each additional parameter (or lag) the numerator increases by one, and the denominator decreases by one.	There *should* be a penalty for being profligate in the parameters. Here's the penalty.

So, when making a decision among models of different lag lengths, choose the model with the smallest FPE.

P. 555: Follow a sequential method in constructing and calculating FPEs:

(1) Start with the variable on the left-hand side. In equation (2.5), this is y_t. Determine the order of y_t's one-dimensional autoregressive process by using the FPE criterion. Suppose the FPE criterion determines the lag to be of order s.

(2) Introduce x_t into the right-hand side of y_t's one-dimensional autoregressive process, of order s, in (1). With y_t's lag fixed at s, let x_t's lag length vary. For each fixed y and varying x combination, choose the pair with smallest FPE. Suppose the lag order of x is n. Thus, x is of order n; y is of order s.

(3) In (1), it is conceivable that y_t picked up the effects of lagged x_t when y_t was expressed as a one-dimensional autoregressive process. So, let the order of x's lag be fixed at n and let the order of y's lag vary from 0 to s. Compute all FPE's. Choose the lag order of y that gives the smallest FPE (conditional on the order of x's lag being n). Suppose y's lag is now m.

(4) Compare the smallest FPE's of steps (1) and (3). If FPE_y (s, 0) in step (1) is less than FPE_y (m, n) in step (3), the one-dimensional autoregressive representation for y is used. Tthis suggests that x does not Granger-cause y. If FPE_y (s, 0) in step (1) is greater than FPE_y (m, n) in step (3), the two-dimensional autoregressive representation for y is used. This suggests that x Granger-causes y.

(5) Repeat steps (1)-(4) with x_t on the left-hand side.

- Results of step (1) are presented in Table 1 for M1, M2, and GNP.

1. FPE of Fitting a One-Dimensional Autoregressive Process for M1, M2, and GNP

The Order of Lags	FPE of M1 x 10^{-4}	FPE of M2 x 10^{-4}	FPE of GNP x 10^{-4}
1	3.775	1.624	2.021
2	3.281	1.610	1.741
3	3.329	1.621	1.786
4	3.003	1.399	1.720
5	2.777	1.399	1.765
6	2.742	1.381	→1.701
7	2.794	1.418	1.741
8	2.671	1.385	1.763
9	→2.572	1.404	1.742
10	2.641	1.409	1.755
11	2.713	1.417	1.792
12	2.743	→1.333	1.816
13	2.802	1.370	1.789
14	2.794	1.394	1.819

· Results of step (2) are presented in Table 2

2. The Optimum Lags of the Manipulated Variable and the FPE of the Controlled Variable

Controlled Variable[a]	Manipulated Variable	The Optimum Lag of Manipulated Variable	FPE $\times 10^{-4}$	
M1 (9)	GNP	4	2.372	⇒ M1 = f [M1(9), GNP(4)]
M2 (12)	GNP	2	1.332	⇒ M2 = f [M2(12), GNP(2)]
GNP (6)	M1	8	1.589	⇒ GNP = f [GNP(6), M1(8)]
GNP (6)	M2	1	1.716	⇒ GNP = f [GNP(6), M2(1)]

[a]The numbers in parentheses indicate the order of autoregressive operator in the controlled variable.

· Note: I see no coverage of step (3) in the paper.

· Results of step (4): Comparison of Table 1 and Table 2 FPEs

	Model	FPE	Source	Granger Causality: Compare FPE Pairs
(1)	M2 = f [M2(12)]	1.333	Table 1	1.333 ≅ 1.332
	M2 = f [M2(12), GNP(2)]	1.332	Table 2	GNP does not Granger-cause M2.
(2)	GNP = f [GNP(6)]	1.701	Table 1	1.701 > 1.589
	GNP = f [GNP(6), M1(8)]	1.589	Table 2	M1 Granger-causes GNP
(3)	GNP = f [GNP(6)]	1.701	Table 1	1.701 < 1.716
	GNP = f [GNP(6), M2(1)]	1.716	Table 2	M2 does not Granger-cause GNP
(4)	M1 = f [M1(9)]	2.572	Table 1	2.572 > 2.372
	M1 = f [M1(9), GNP(4)]	2.372	Table 2	GNP Granger-causes M1

COINTEGRATION

A stationary variable will tend to wander extensively. (This is what makes it nonstationary). But some *pairs* of nonstationary variables can be expected to wander in such a way that they do not drift too far apart. (Examples of such variables are short- and long-term interest rates, prices and wages, household income and expenditures, imports and exports, spot and future prices of a commodity, and exchange rates determined in different markets.)

Such variables are said to be *cointegrated*:

- Although individually they are I(1), a particular linear combination of them is I(0). (Economists have an interest in cointegration; it provides a formal framework for testing and estimating long-run (equilibrium) relationships among economic variables).

Recall what Granger and Newbold concluded (1974) regarding nonstationary variables in a regression model (Enders pages 216-218).

- Granger and Newbold warned of regression one I(1) variable on another I(1) variable and called these regressions "spurious", since the least squares estimator breaks down in this case.

Granger identified a situation when the regression of an I(1) process on another I(1) was *not spurious*. In a situation where the variables are cointegrated, the least squares estimator works better because it converges to the true parameter value faster than usual.

Conclusion:

Differencing is not the only means of eliminating unit roots. So, if data are found to have unit roots -- before differencing -- test for cointegration.

Q: How do we test for cointegration?

A: Although individually two variables may be I(1), a particular linear combination of them is I(0). This *is* COINTEGRATION. Consider our original, "problematic" equation:

$$y_t = \beta_1 + \beta_2 x_2 + \varepsilon_t.$$

What is a good candidate for the magical linear combination that leads to stationarity? Try:

$$\varepsilon_t = y_t - \beta_1 - \beta_2 x_t$$

Recall that y_t and x_t "drift along" together. It may be reasonable to expect that $\varepsilon_t \sim (0, \sigma^2)$. Note: y_t and x_t are I(1), but ε_t may very well be I(0).

We can now explore the behavior of the $\underline{\hat{\varepsilon}}_t$ vector to see if it follows a unit root process. How?

- Treat $\underline{\hat{\varepsilon}}_t$ as we would $\underline{y}_t$ or $\underline{x}_t$.
- Use TSP, pages 16-10 to 16-13.
  - UROOT (C,0) Y X
  - This is equivalent to LS Y C X, and it retains coefficients and residuals.
  - UROOT then runs LS D(RESID) RESID(-1)

How to proceed:

- Consider the first order autoregressive model of the residuals

$$\hat{\varepsilon}_t = \rho \hat{\varepsilon}_{t-1} + \upsilon_t \rightarrow \text{what does this look like?}$$

- In light of what we did with unit root tests of individual series, $\hat{\varepsilon}_t$ is stationary if $|\rho|<1$. But if $\rho=1$, the errors are nonstatioinary.
- To carry out the test under $H_0 : \rho=1$
  - Subtract $\hat{\varepsilon}_{t-1}$ from both sides of the equation

$$\hat{\varepsilon}_t - \hat{\varepsilon}_{t-1} = \rho \hat{\varepsilon}_{t-1} - \hat{\varepsilon}_{t-1} + \upsilon_t$$

$$\Delta\hat{\varepsilon}_t = (\rho-1)\hat{\varepsilon}_{t-1} + \upsilon_t = \gamma \hat{\varepsilon}_{t-1} + \upsilon_t \text{ (where } \gamma = \rho-1)$$

  - The null hypothesis is therefore H_0: $\rho = 1$ or $\gamma = 0$
  - The null hypothesis is rejected, on the basis of a one-tailed t-test, if $t<t_c^*$ where t_c^* is a critical value.
- Include an intercept, trend, and lagged differences of the errors (ADF) components.
- Known as the Engle-Granger test.
- Use TSP:

  UROOT (C, 1) Y X , which is equivalent to

  LS Y C X and LS D(RESID) RESID(-1) D(RESID(-1))

Mechanics Of Testing For Cointegration

- $Y_t = \beta_0 + \beta_1 X_t + \varepsilon_t$
  $\uparrow$ I(1)? $\uparrow$ I(1)?

- → Test to see if Y_t is I(1) by using UROOT (,)Y.
  → Test to see if X_t is I(1) by using UROOT (,)X.
  → If both are I(1), proceed with cointegration test.
  → If both are not I(1), *do not* pursue cointegration.
  If both are not I(1), they *can't* be cointegrated.

- Estimate $Y_t = \beta_0 + \beta_1 X_t + \varepsilon_t$: Ordinary Least Squares
  → Output $\hat{\beta}_0$, $\hat{\beta}_1$, $\hat{\varepsilon}_t$. Call $\hat{\varepsilon}_t$ "RESID"
  → Try to find that linear combination of two I(1) variables that results in an I(0) variable.
  → A possibility: $\varepsilon_t = Y_t - \beta_0 - \beta_1 X_t$
  → To see if ε_t is I(0), do UROOT on $\hat{\varepsilon}_t$ (RESID)

- TSP: pages 16-10 to 16-13
  UROOT (C,0) Y X
  This is equivalent to doing:
  LS Y C X
  LS D(RESID) RESID(-1)

UROOT (C,1) Y X →LS D(RESID) RESID(-1) D(RESID(-1))

ERROR CORRECTION MODEL

1. (a) Cointegration means that, in the long run, we can be assured that the two series move together. We would like to account for disturbances in the short run that will keep us on track in terms of one-step-ahead forecasts.

   (b) The error correction model (ECM) is a way of accounting for short run dynamics *when* we can be assured that a long run equilibrium exists.

      (i) A simple specification:

$$\Delta y_t = \alpha_0 + \alpha_1 \Delta x_t + \alpha_2 (y_{t-1} - \beta_1 - \beta_2 x_{t-1}) + \upsilon_t$$

contemporaneous change in y due to a change in x in this cointegrated system	an error from the previous period that compensates for dynamic changes in the short run.

      (ii) This specification is called the *error correction* model (ECM).

      (iii) Notice that it is non linear in the parameters; check out how α_2, β_1, and β_2 are related.

      (iv) Notice that $\Delta y_t = f(\Delta x_t) \Leftrightarrow y_t - y_{t-1} = f(x_t - x_{t-1})$. This says that, in order to forecast y for anything beyond one-step ahead, we must forecast x.

4. TESTING FOR COINTEGRATION: THE ENGLE–GRANGER METHODOLOGY

To explain the Engle–Granger testing procedure, let us begin with the type of problem likely to be encountered in applied studies. Suppose that two variables—say, y_t and z_t—are believed to be integrated of order 1 and we want to determine whether there exists an equilibrium relationship between the two. Engle and Granger (1987)

propose a straightforward test whether two $I(1)$ variables are cointegrated of order $CI(1, 1)$.

STEP 1: Pretest the variables for their order of integration. By definition, cointegration necessitates that the variables be integrated of the same order. Thus, the first step in the analysis is to pretest each variable to determine its order of integration. The Dickey–Fuller, augmented Dickey–Fuller, and/or Phillips–Perron tests discussed in Chapter 4 can be used to infer the number of unit roots (if any) in each of the variables. If both variables are stationary, it is not necessary to proceed since standard time-series methods apply to stationary variables. If the variables are integrated of different orders, it is possible to conclude that they *are* not cointegrated.[10]

STEP 2: Estimate the long-run equilibrium relationship. If the results of Step 1 indicate that both $\{y_t\}$ and $\{z_t\}$ are $I(1)$, the next step is to estimate the long-run equilibrium relationship in the form:

$$y_t = \beta_0 + \beta_1 z_t + e_t \tag{6.30}$$

If the variables are cointegrated, an OLS regression yields a "super-consistent" estimator of the cointegrating parameters β_0 and β_1. Stock (1987) proves that the OLS estimates of β_0 and β_1 converge faster than in OLS models using stationary variables. To explain, reexamine the scatter plot shown in Figure 6.1. You can see that the effect of the common trend dominates the effect of the stationary component; both variables seem to rise and fall in tandem. Hence, there is a strong linear relationship as shown by the regression line drawn in the figure.

In order to determine if the variables are actually cointegrated, denote the residual sequence from this equation by $\{\hat{e}_t\}$. Thus, $\{\hat{e}_t\}$ is the series of the estimated residuals of the long-run relationship. If these deviations from long-run equilibrium are found to be stationary, the $\{y_t\}$ and $\{z_t\}$ sequences are cointegrated of order (1, 1). It would be convenient if we could perform a Dickey–Fuller test on these residuals to determine their order of integration. Consider the autoregression of the residuals:

$$\Delta\hat{e}_t = a_1 \hat{e}_{t-1} + \epsilon_t \tag{6.31}$$

Since the $\{\hat{e}_t\}$ sequence is a residual from a regression equation, there is no need to include an intercept term; the parameter of interest in (6.31) is a_1. If we cannot reject the null hypothesis $a_1 = 0$, we can conclude that the residual series contains a unit root. Hence, we conclude that the $\{y_t\}$ and $\{z_t\}$ sequences are *not* cointegrated. The more precise wording is awkward because of a triple negative, but to be technically correct, *if it is not possible to reject the null hypothesis $|a_1| = 0$, we cannot reject the hypothesis*

that the variables are not cointegrated. Instead, the rejection of the null hypothesis implies that the residual sequence is stationary.[11] Given that both $\{y_t\}$ and $\{z_t\}$ were found to be $I(1)$ and the residuals are stationary, we can conclude that the series are cointegrated of order (1, 1).

In most applied studies, it is not possible to use the Dickey–Fuller tables themselves. The problem is that the $\{\hat{e}_t\}$ sequence is generated from a regression equation; the researcher does not know the actual error $\hat{e}_t$, only the estimate of the error $\hat{e}_t$. The methodology of fitting the regression in (6.30) selects values of β_0 and β_1 that minimize the sum of squared residuals. Since the residual variance is made as small as possible, the procedure is prejudiced toward finding a stationary error process in (6.31). Hence, the test statistic used to test the magnitude of a_1 must reflect this fact. Only if β_0 and β_1 were known in advance and used to construct the true $\{e_t\}$ sequence would an ordinary Dickey–Fuller table be appropriate. Fortunately, Engle and Granger provide test statistics that can be used to test the hypothesis $a_1 = 0$. When more than two variables appear in the equilibrium relationship, the appropriate tables are provided by Engle and Yoo (1987).

If the residuals of (6.31) do not appear to be white-noise, an augmented Dickey–Fuller test can be used instead of (6.31). Suppose that diagnostic checks indicate that the $\{\epsilon_t\}$ sequence of (6.31) exhibits serial correlation. Instead of using the results from (6.31), estimate the autoregression:

$$\Delta\hat{e}_t = a_1\hat{e}_{t-1} + \sum_{i=1}^{n} a_{i+1}\Delta\hat{e}_{t-i} + \epsilon_t \qquad (6.32)$$

Again, if $-2 < a_1 < 0$, we can conclude that the residual sequence is stationary and $\{y_t\}$ and $\{z_t\}$ are $CI(1, 1)$.

STEP 3: Estimate the error-correction model. If the variables are cointegrated (i.e., if the null hypothesis of no cointegration is rejected), the residuals from the equilibrium regression can be used to estimate the error-correction model. If $\{y_t\}$ and $\{z_t\}$ are $CI(1, 1)$, the variables have the error-correction form:

$$\Delta y_t = \alpha_1 + \alpha_y(y_{t-1} - \beta_1 z_{t-1}) + \sum_{i=1} \alpha_{11}(i)\Delta y_{t-i} + \sum_{i=1} \alpha_{12}(i)\Delta z_{t-i} + \epsilon_{yt} \qquad (6.33)$$

$$\Delta z_t = \alpha_2 + \alpha_z(y_{t-1} - \beta_1 z_{t-1}) + \sum_{i=1} \alpha_{21}(i)\Delta y_{t-i} + \sum_{i=1} \alpha_{22}(i)\Delta z_{t-i} + \epsilon_{zt} \qquad (6.34)$$

where β_1 = the parameter of the cointegrating vector given by (6.30)

ϵ_{yt} and ϵ_{zt} = white-noise disturbances (which may be correlated with each other)

and α_1, α_2, α_y, α_z, $\alpha_{11}(i)$, $\alpha_{12}(i)$, $\alpha_{21}(i)$, and $\alpha_{22}(i)$ are all parameters.

Engle and Granger (1987) propose a clever way to circumvent the cross-equation restrictions involved in the direct estimation of (6.33) and (6.34). The value of the residual $\hat{e}_{t-1}$ estimates the deviation from long-run equilibrium in period $(t - 1)$. Hence, it is possible to use the saved residuals $\{\hat{e}_{t-1}\}$ obtained in Step 2 as an instrument for the expression $y_{t-1} - \beta_1 z_{t-1}$ in (6.33) and (6.34). Thus, using the saved residuals from the estimation of the long-run equilibrium relationship, we can estimate the error-correcting model as

$$\Delta y_t = \alpha_1 + \alpha_y \hat{e}_{t-1} + \sum_{i=1} \alpha_{11}(i)\Delta y_{t-i} + \sum_{i=1} \alpha_{12}(i)\Delta z_{t-i} + \epsilon_{yt} \tag{6.35}$$

$$\Delta z_t = \alpha_2 + \alpha_z \hat{e}_{t-1} + \sum_{i=1} \alpha_{21}(i)\Delta y_{t-i} + \sum_{i=1} \alpha_{22}(i)\Delta z_{t-i} + \epsilon_{zt} \tag{6.36}$$

Other than the error-correction term $\hat{e}_{t-1}$, $\hat{e}$ Equations (6.35) and (6.36) constitute VAR in first differences. This *near VAR* can be estimated using the same methodology developed in Chapter 5. All the procedures developed for a VAR apply to the near VAR. Notably:

1. OLS is an efficient estimation strategy since each equation contains the same set of regressors.
2. Since all terms in (6.35) and (6.36) are stationary [i.e., Δy_t and its lags, Δz_t and its lags, and $\hat{e}_{t-1}$ are $I(0)$], the test statistics used in traditional VAR analysis are appropriate for (6.35) and (6.36). For example, lag lengths can be determined using a χ^2 test and the restriction that all $\alpha_{jk}(i) = 0$ can be checked using an F-test. If there is a single cointegrating vector, restrictions concerning α_y or α_z can be conducted using a t-test. Asymptotic theory indicates α_y and α_z converge to a t-distribution as sample size increases.

STEP 4: Assess model adequacy. There are several procedures that can help determine whether the estimated error-correction model is appropriate.

1. You should be careful to assess the adequacy of the model by performing diagnostic checks to determine whether the residuals of the near

Link Between VAR And Where We Are Now (Error Correction)

Start with a structural VAR (p. 294, equation 5.19):

$$y_t = \beta_0 + \beta_1 x_1 + \beta_2 y_{t-1} + \beta_3 x_{t-1} + \varepsilon_t \qquad (1)$$

Develop (1) into a VAR in standard form (p. 295, equation 5.22a)

$$y_t = \alpha_0 + \alpha_1 y_{t-1} + \alpha_2 x_{t-1} + \mu_t \qquad (2)$$

This equation is nice because it allows *one*-step-ahead forecasts using actual, rather than forecasted, values of y_{t-1} and x_{t-1}. Of course, *two*-or-more-step ahead forecasts promote increased forecast error due to y_{t-1} and x_{t-1} being unknown at the time of the forecast.

Q: Is there anything we can do to lessen the impact of such an error? Is there a correction that will put us back on target as we make these longer-range forecasts?

A: Yes:

We do this by

(a) Using the structural VAR in (1).

(b) Imposing restrictions, which reflect the belief that, in the long run, all y's settle down to a value, say, y_t and x's to say, x_t -- their respective equilibrium values.

(c) The result of (a) and (b) is an error correction model.

Start with (a)

- $y_t = \beta_0 + \beta_1 x_t + \beta_2 y_{t-1} + \beta_3 x_{t-1} + \varepsilon_t$
- Set y_{t-1} to y_t and x_{t-1} to x_t, suggesting that, in the long run, an equilibrium is achieved.
- Another way of stating the above: what relationship must exist among the β_is so that y_t-x_t is a constant?
- Take these statements apart:

$$y_t = \beta_0 + \beta_1 x_t + \beta_2 \underset{\downarrow}{y_{t-1}} + \beta_3 \underset{\downarrow}{x_{t-1}} + \varepsilon_t$$

$y_t = \beta_0 + \beta_1 x_t + \beta_2 y_t + \beta_3 x_t + \varepsilon_t$: setting $y_t = y_{t-1}$; $x_t = x_{t-1}$

$\beta_0 = y_t - \beta_2 y_t - \beta_1 x_t - \beta_3 x_t$: isolating the constant and setting $\varepsilon_t = 0$ in the long run.

$\beta_0 = (1 - \beta_2) y_t - (\beta_1 + \beta_3) x_t$: gathering terms

$(1 - \beta_2) = (\beta_1 + \beta_3)$ or $\beta_1 + \beta_2 + \beta_3 = 1$: imposing the restriction that, in the long run, y and x will grow at the same rate, so that in equilibrium (y-x) will be a constant

Summary to now:

- Given an equation in a structural VAR (or in a primitive system)

$$y_t = \beta_0 + \beta_1 x_t + \beta_2 y_{t-1} + \beta_3 x_{t-1} + \varepsilon_t \quad .$$

  Imposing the restriction $\beta_1 + \beta_2 + \beta_3 = 1$ results in a special case of this original equation:

$$y_t = \beta_0 + \beta_1 x_t + \beta_2 y_{t-1} + \beta_3 x_{t-1} + \varepsilon_t$$

$$\text{s.t.} \quad \beta_1 + \beta_2 + \beta_3 = 1$$

  This special case imposes long run equilibrium.

  This special case is called the error correction model.

- We could test this restriction straightforwardly in TSP.

Putting The Original Equation Into The Engle-Granger Form

- Subtract y_{t-1} from each side
- Add and subtract $\beta_1 x_{t-1}$ to and from the right-hand side
- Impose the restriction

$y_t = \beta_0 + \beta_1 x_t + \beta_2 y_{t-1} + \beta_3 x_{t-1} + \varepsilon_t$: original equation

$y_t - y_{t-1} = \beta_0 + \beta_1 x_t + \beta_2 y_{t-1} - y_{t-1} + \beta_3 x_{t-1} + \varepsilon_t$: subtract y_{t-1} from each side

$y_t - y_{t-1} = \beta_0 + \beta_1 x_t - \beta_1 x_{t-1} + \beta_2 y_{t-1} - y_{t-1} + \beta_3 x_{t-1} + \beta_1 x_{t-1} + \varepsilon_t$: subtract and add $\beta_1 x_{t-1}$

$y_t - y_{t-1} = \beta_0 + \beta_1 (x_t - x_{t-1}) + (\beta_2 - 1) y_{t-1} + (\beta_3 + \beta_1) x_{t-1} + \varepsilon_t$: collect terms

$y_t - y_{t-1} = \beta_0 + \beta_1 (x_t - x_{t-1}) + (\beta_2 - 1) y_{t-1} - (-\beta_3 - \beta_1) x_{t-1} + \varepsilon_t$: changing signs on the last term

$y_t - y_{t-1} = \beta_0 + \beta_1 (x_t - x_{t-1}) + \gamma \; y_{t-1} - \gamma \; x_{t-1} + \varepsilon_t$: forcing β_2-1 = -β_3-β_1 calling each side γ)

$y_t - y_{t-1} = \beta_0 + \beta_1 (x_t - x_{t-1}) + \gamma (y_{t-1} - x_{t-1}) + \varepsilon_t$: collect terms

$\Delta y_t = \beta_0 + \beta_1 \Delta x_t + \gamma (y_{t-1} - x_{t-1}) + \varepsilon_t$: writing in Δ notation

or

Engle and Granger propose estimating $y_t = \alpha_1 + \alpha_2 x_t + \varepsilon$ by OLS, using $\hat{\alpha}_1$ and $\hat{\alpha}_2$ to form $e_t = y_t - \hat{\alpha}_1 - \hat{\alpha}_2 x_t$, lagging the e_t series, and using this e_{t-1} series at this point. (Important note: a stationarity test must be done on e_t prior to this step.

$\Delta y_t = \beta_0 + \beta_1 \Delta x_t + \gamma e_{t-1} + \varepsilon_t$

Using A VAR In Standard Form Rather Than In Structural Form To Get The ECM

$y_t = \alpha_0 + \alpha_1 y_{t-1} + \alpha_2 x_{t-1} + \mu_t$:	VAR in standard form
$y_t - y_{t-1} = \alpha_0 + \alpha_1 y_{t-1} - y_{t-1} + \alpha_2 x_{t-1} + \mu_t$:	subtract y_{t-1} from each side
$y_t - y_{t-1} = \alpha_0 + (\alpha_1 - 1) y_{t-1} + \alpha_2 x_{t-1} + \mu_t$:	collect terms
$y_t - y_{t-1} = \alpha_0 + (\alpha_1 - 1) y_{t-1} + (-\alpha_2)(-x_{t-1}) + \mu_t$:	changing signs on the last term
$y_t - y_{t-1} = \alpha_0 + \gamma \; y_{t-1} - \gamma \; x_{t-1} + \mu_t$:	forcing $\alpha_1 - 1 = -\alpha_2$ (call each side γ)
$y_t - y_{t-1} = \alpha_0 + \gamma \; (y_{t-1} - x_{t-1}) + \mu_t$:	collect terms
$\Delta y_t = \alpha_0 + \gamma \; (y_{t-1} - x_{t-1}) + \mu_t$:	writing in Δ notation

Replace using Engle-Granger suggestion

$\Delta y_t = \alpha_0 + \gamma \; e_{t-1} + \mu_t \quad \Rightarrow$ Note: By starting with a VAR in *standard* form (rather than in the *structural* or *primitive* form), the resulting error correction model will have *no* contemporaneous right-hand-side variables.

This is incredibly useful when doing forecasts. (Compare this form of the ECM with the ECM developed from the *structural* VAR, which is: $\Delta y_t = \beta_0 + \beta_1 \Delta x_t + \gamma e_{t-1} + \varepsilon_t$)

- Why do you think this is called an Error Correction Model (ECM)?

How do you do post-sample forecasts?

(1) For one-step ahead, you have e_{t-1}; it does not have to be forecasted.

(2) (1) is plugged into the within-sample equation $\Delta y_t = \hat{\alpha}_0 + \hat{\gamma} e_{t-1}$. This result is $\widehat{\Delta y_t}$.

(3) We need a forecast (one-step-ahead) for y_t. Make use of your result in (2):

$$\widehat{\Delta y_t} = y_t - y_{t-1}$$

you have!   you need   you have!

(4) Forecast of y_t: $\hat{y}_t = \widehat{\Delta y_t} + y_{t-1}$

Augmenting The Previous Standard VAR Form With More Lags To Get A More Complicated ECM

(1) $y_t = \alpha_0 + \alpha_1 y_{t-1} + \alpha_2 y_{t-2} + \alpha_3 x_{t-1} + \alpha_4 x_t + \mu_t$: VAR in standard form

(2) $y_t - y_{t-1} = \alpha_0 + \alpha_1 y_{t-1} + \alpha_2 y_{t-2} + \alpha_3 x_{t-1} + \alpha_4 x_{t-2} + \mu_t$: subtract y_{t-1} from each side

(3) $y_t - y_{t-1} = \alpha_0 + \alpha_1 y_{t-1} - y_{t-1} + \alpha_2 y_{t-1} - \alpha_2 y_{t-1} + \alpha_2 y_{t-2} + \alpha_3 x_{t-1} + \alpha_4 x_{t-2} + \mu_t$: add and subtract $\alpha_2 y_{t-1}$ to and from the right side

(4) $y_t - y_{t-1} = \alpha_0 + [\alpha_1 y_{t-1} - y_{t-1} + \alpha_2 y_{t-1}] - \alpha_2 [y_{t-1} - y_{t-2}] + \alpha_3 x_{t-1} + \alpha_4 x_{t-2} + \mu_t$: bracketing y terms and factoring out an α_2

(5) $y_t - y_{t-1} = \alpha_0 + [\alpha_1 y_{t-1} - y_{t-1} + \alpha_2 y_{t-1}] - \alpha_2 [y_{t-1} - y_{t-2}] + \alpha_3 x_{t-1} + \alpha_4 x_{t-1} - \alpha_4 x_{t-1} + \alpha_4 x_{t-2} + \mu_t$: add and subtract $\alpha_4 x_{t-1}$ to and from the right side

(6) $y_t - y_{t-1} = \alpha_0 + [\alpha_1 y_{t-1} - y_{t-1} + \alpha_2 y_{t-1}] - \alpha_2 [y_{t-1} - y_{t-2}] + [\alpha_3 x_{t-1} + \alpha_4 x_{t-1}] - \alpha_4 [x_{t-1} - x_{t-2}] + \mu_t$: bracketing x terms and factoring out an α_4

(7) $y_t - y_{t-1} = \alpha_0 + [\alpha_1 + \alpha_2 - 1] y_{t-1} - \alpha_2 [y_{t-1} - y_{t-2}] + [\alpha_3 + \alpha_4] x_{t-1} - \alpha_4 [x_{t-1} - x_{t-2}] + \mu_t$: factoring out y_{t-1} and x_{t-1}

(8) $\Delta y_t = \alpha_0 + [\alpha_1 + \alpha_2 - 1] y_{t-1} - \alpha_2 \Delta y_{t-1} + [\alpha_3 + \alpha_4] x_{t-1} - \alpha_4 \Delta x_{t-1} + \mu_t$: writing in Δ notation

(9) $\Delta y_t = \alpha_0 + [\alpha_1 + \alpha_2 - 1] y_{t-1} - \alpha_2 \Delta y_{t-1} + [-\alpha_3 - \alpha_4] [-x_{t-1}] - \alpha_4 \Delta x_{t-1} + \mu_t$: changing signs on all items in one term

(10) $\Delta y_t = \alpha_0 + \quad \gamma y_{t-1} \quad - \alpha_2 \Delta y_{t-1} + \quad \gamma \quad [-x_{t-1}] - \alpha_4 \Delta x_{t-1} + \mu_t$: restricting coefficients on lagged levels to be the same

(11) $\Delta y_t = \alpha_0 + \gamma (y_{t-1} - x_{t-1}) - \alpha_2 \Delta y_{t-1} - \alpha_4 \Delta x_{t-1} + \mu_t$
Replace using Engle-Granger suggestion

(12) $\Delta y_t = \alpha_0 + \gamma e_{t-1} - \alpha_2 \Delta y_{t-1} - \alpha_4 \Delta x_{t-1} + \mu_t$

- This form *imposes* the restriction that $\alpha_1 + \alpha_2 - 1 = \alpha_3 - \alpha_4 \Leftrightarrow \alpha_1 + \alpha_2 + \alpha_3 + \alpha_4 = 1$
- You could test the validity of this restriction.
- If "error correction" is valid, Δy_t again is useful for "correcting" forecasts.
- Note: Equation (12) is comparable to Enders, p. 376, equation 6.35 or equation 6.36.