Lecture Notes 3: Single Equation Cointegration

Carl Bonham, Ph.D., August, 2005

Department of Economics, University of Hawaii
and University of Hawaii Economic Research Organization
1 Co-integration Conceptual Issues and Testing

So far we know that economic series tend to behave as if they were generated by random walk processes. In other words they have some form of long run trend, probably stochastic. Early econometricians such as Box and Jenkins (1960) suggested that economists should difference their data before estimating ARMA models because their methods (TSA) relied on classical statistics and hence stationary series. However, differencing series before analysis throws out the long-run information present in the data and makes it less obvious how to model the long-run equilibrium relationships suggested by economic theory. The desire to evaluate both the short-run and long-run characteristics of economic relationships led Davidson, Hendry, Serba, and Yeo (1978) to work with models in their error correction form. It turns out, that series that are jointly generated by an error correction model are also both integrated and co-integrated. Before discussing how to test for co-integration and set up error correction models, let’s consider the concept of equilibrium in more detail.

Consider a pair of time series, \( x_t \) and \( y_t \). There may be some reason the economy ensures that points in the scatter of \( x \) and \( y \) lie near some line rather than elsewhere in the scatter. The reason may be some government policy, the decision making of consumers or business persons, or some other market function. For example, arbitrage insures that prices on the same commodity in two different regions of a country tend to force prices to be the same (after adjusting for transportation and storage). In two different countries, PPP should lead to a long-run equilibrium relationship between tradeable goods prices and exchange rates or between asset yields and exchange rates.

Now consider a scatter plot of these two series \( x \) and \( y \). A line through a scatter of points will act as an attractor if some mechanism exists which ensures that a combination
of the two series cannot drift too far from the line. Although shocks may occur which move the economy away from the line, there will be an overall tendency to move toward the line. We can think of the line as an 'equilibrium' such as a center of gravity for the economic relationship under consideration.

The concept of equilibrium that we are interested in here is a long-run concept. A long-run equilibrium relationship implies a systematic co-movement of economic variables. Crudely put, we can define an equilibrium relationship between two variables \( y \) and \( x \), \( f(y, x) = 0 \), if the amount \( \epsilon_t \equiv f(y_t, x_t) \) by which the two variables deviate from the equilibrium is a mean-zero stationary process. In other words, the “error”, or the discrepancy between the observed relationship between \( y \) and \( x \) and the hypothesized equilibrium relationship has a fixed distribution, centered on zero, that does not change over time. Thus, the equilibrium error cannot grow without bound. If it could, then the economic relationship is not an equilibrium relationship because the two variables are free to move further and further away from that relationship as time passes.

**Trivial Cointegration**

The above definition of an equilibrium relationship will hold automatically when we consider two series that are both stationary. Remember that a linear combination of any two stationary series will be stationary, regardless as to whether a long run equilibrium relationship exists between the two variables. That is, for \( y_t \sim I(0) \), and \( x_t \sim I(0) \), any \( f(y_t, x_t) = (y_t - \beta x_t) = \epsilon_t \sim I(0) \). Whether or not there exists a non-zero \( \beta \) which describes a true long-run equilibrium relationship (corresponding to a nonzero derivative between \( y_t \) and \( x_t \)), any arbitrary value of \( \beta \) will satisfy the statistical definition of equilibrium given above. Thus, the concept of statistical equilibrium defined above is most useful when we are examining relationships between economic time series which grow over time, i.e. macro
time series. In this case, if the true equilibrium relationship is \( y_t = \gamma x_t \), then \( y_t - \beta x_t \) will be \( I(1) \) for any \( \beta \neq \gamma \). For any parameter other than the true equilibrium parameter \( \gamma \), the discrepancy \( y_t - \beta x_t \) will be \( I(1) \) since it deviates from the true equilibrium error by the amount \( (\beta - \gamma)x_t \). With bivariate relationships, only the true equilibrium relationship can produce a stationary error. On the other hand, when we work with more than two variables, there may be more than one equilibrium relation. Below we explore bivariate relationships thoroughly, in later lectures we will study multivariate co-integrating relationships.

1.1 Balancing Equations

Before we get further into the issues of testing for and analyzing co-integrating relationships, let’s review the issue of balancing regression equations. Econometricians have been aware of the importance of balance since at least Box and Jenkins (1976). As an example, consider only the issue of balance in mean. A constant-mean regressor cannot fully explain a variable with a linear trend, and a variable with a linear trend cannot explain a series with an exponential trend. A common practice to balance regression equations is to take logarithms (see Ermini and Hendry (1991)).

Nonstationarities such as generated by \( I(1) \) processes can produce inbalance, in both mean and variance, in regression equations if the researcher is not particularly careful. In general, an \( I(d-1) \) process cannot adequately explain the behavior of an \( I(d) \) process. Thus, if \( d = 1 \), we know that an \( I(0) \) process (which has a constant mean and variance) cannot possibly explain all of the behavior of an \( I(1) \) process whose mean and variance trend upward overtime.
In addition, although two processes are both \( I(d) \), it does not necessarily follow that one can adequately explain (balance) the other. In fact, regressions with \( I(d) \) processes will only balance if co-integration occurs, and otherwise produce spurious regressions.

1.2 Guide to Building Bivariate Systems

Consider a bivariate system relating \( y_t \) and \( x_t \). A simple recommended procedure is to:

1. Perform unit-root (or seasonal unit-root) tests on both variables, to determine their orders of integration. (For convenience, let’s assume that both series are \( I(1) \). Experience tells us that most real macro series are at most \( I(1) \)).

2. If both variables are found to be \( I(0) \), proceed to your model identification stage, and choose your model in levels or log levels.

3. If one variable is \( I(0) \) and the other \( I(1) \), the later must be differenced to balance your model. Next proceed as in step 2 above. It is also possible to include multiple lags of an integrated right hand side variable to balance the equation. But this complicates inference.

4. If both variables are \( I(1) \), there are two possibilities:
   - The two variables are not co-integrated and no long run equilibrium relationship exists between them. In this case, models estimated in levels will result in \( I(1) \) residuals and spurious results that may look appealing but will be subject to structural instability and lack of interpretability.
   - The two variables are co-integrated. In this case, the model may be written in error correction form to allow for standard inference regarding short run impact effects of \( x \) on \( y \), disequilibrium feedback, and long run equilibrium response.

The distinction between modeling \( I(0) \) and \( I(1) \) processes entails three possible model structures: a model in levels if the processes are \( I(0) \), a model in first differences if the series are \( I(1) \) but not co-integrated, and an error correction model if the variables are \( I(1) \) and co-integrated.
1.3 Definition of Co-integration

If $x_t, y_t$ are $I(1)$ but there exists a linear combination

$$z_t = m + ax_t + by_t$$

which is both $I(0)$ and has a zero mean, then $x_t, y_t$ are said to be co-integrated. More generally, time series $x_t$ and $y_t$ are said to be co-integrated of order $d, b$ where $d \geq b \geq 0$, if (1) both series are integrated of order $d$, and (2) there exists a linear combination of these variables, say $\alpha_1 x_t + \alpha_2 y_t$, which is integrated of order $d - b$. The vector $[\alpha_1, \alpha_2]$ is called the co-integrating vector, and the co-integrating relationship is written as: $x_t, y_t \sim CI(d, b)$.

It is also possible to consider a multivariate co-integrating relationship where $x$ is an $n \times 1$ vector of $I(d)$ series, and there exists some linear combination $x' \alpha \sim I(d - b)$, where $\alpha$ is an $n \times 1$ co-integrating vector. In the case of more than two series ($n > 2$), it is possible to have each element of the $x$ vector be integrated of different order.

Generally, we are interested in the case where the individual series are $I(1)$, and the linear combination is stationary.
1.4 ECM Example

To see how co-integration might occur, suppose we have two time series such as the yield on two-year, $x_t$, and one-year, $y_t$, government bonds. Consider the following construction

\[
\begin{align*}
x_t &= aw_t + \epsilon_{xt} \\
y_t &= w_t + \epsilon_{yt} \\
w_t &= w_{t-1} + v_t
\end{align*}
\]

In this system, $w_t$ and therefore both $x_t$ and $y_t$ are $I(1)$, yet $\epsilon_{xt}, \epsilon_{yt}$, are both $I(0)$ with zero means. Generally we expect that a linear combination of $I(1)$ variables will also be $I(1)$. However,

\[
z_t = x_t - ay_t \sim I(0)
\]

is stationary, and has a zero mean. Thus, we say that $x_t$ and $y_t$ co-integrate with co-integrating parameter $a$. These two series co-integrate because their $I(1)$ behavior arises from the same source. In other words they share a common $I(1)$ factor $w_t$. In fact, “a pair of co-integrated series will always have an $I(1)$ common factor representation.” (Granger, 1996)
1.4.1 Uniqueness

Note that in the bivariate case, the cointegrating parameter $a$ is unique. To see this, suppose that we have

$$v_t = x_t - \gamma y_t \sim I(0). \quad (4)$$

Subtracting (4) from (3), we have

$$z_t - v_t = (a - \gamma) y_t. \quad (5)$$

Under the assumption that both $a$ and $\gamma$ are cointegrating parameters, the LHS of (5) must be $I(0)$, and this is only possible if $a = \gamma$.

1.4.2 Different Representations

From the ‘statistical’ equilibrium defined above, it is clear that the line $x = ay$ represents an attractor for these two series. Consider the scatter plot of the two series. Since $x_t, y_t$ are both $I(1)$, the points $(x_t, y_t)$ will move widely around the $x - y$ plane. However, $z_t = x_t - ay_t \sim I(0)$, with zero mean, thus there will be a tendency for the points in the scatter to collect around the line, and this line acts as an attractor.

A sufficient condition for the existence of an attractor line is that the two series be co-integrated. Furthermore, the Granger representation theorem (Engle and Granger (1991, pp. 7-8)) proves that when two series are co-integrated, they are also generated by a model
referred to as an error correction model. The converse is also true, that is, series which are generated by an ECM must be co-integrated.\(^1\) To see this, write the first equation in 2 as

\[ x_t = ay_t + \epsilon_t \quad \epsilon_t = \rho\epsilon_{t-1} + u_t \quad \text{where } |\rho| < 1, \tag{6} \]

and using the “equilibrium error” \( z_t = x_t - ay_t = \epsilon_t \) we can write,

\begin{align*}
  z_t &= \rho z_{t-1} + u_t \\
  \Delta z_t &= (\rho - 1)z_{t-1} + u_t \\
  \text{replacing } z_t \text{ with } x_t - ay_t \\
  \Delta x_t &= \Delta y_t + (\rho - 1)z_{t-1} + u_t \\
  \tag{7}
\end{align*}

In other words, the cointegrating regression may be rewritten in error correction form. The “equilibrium error”, \( z_t \) enters lagged one period with a loading parameter \((\rho - 1)\) that is less than zero.

More generally, both of \( x_t, y_t \) can be written in error correction form:

\begin{align*}
  \Delta y_t &= m_1 + \rho_1 z_{t-1} + lags(\Delta y_t, \Delta y_t) + \epsilon_{yt} \\
  \Delta x_t &= m_2 + \rho_2 z_{t-1} + lags(\Delta x_t, \Delta y_t) + \epsilon_{xt} \\
  \tag{8}
\end{align*}

\(^1\)The Granger Representation Theorem actually proves that a co-integrated system of variables can be written in three isomorphic forms: the vector autoregressive (VAR), an error-correction, and moving average forms.
where \((\epsilon_{yt}, \epsilon_{xt})\) is a bivariate white noise, \(z_t = x_t - ay_t\), and \textit{at least one of the }\rho_1, \rho_2 \textit{is different from zero}. Notice that if \(x, y\) are co-integrated, then each of the components in (8) are \(I(0)\), so we have balanced equations.

In contrast, if \(x_t, y_t\) are \(I(1)\) but do not CI, then \(z_t \sim I(1)\), and cannot properly explain the behavior of the short-memory processes \((\Delta x_t, \Delta y_t)\). In this case, the equations can only balance if both \(\rho_1, \rho_2\) are equal to zero. \textit{Co-integration is a necessary condition for equation (8) to hold, and vise versa.}

### 1.4.3 Equilibrium Error

The linear attractor defines a statistical equilibrium where by \(z_t \sim I(0)\). \(z_t\) measures the extent to which the system deviates from its equilibrium and is therefore often referred to as the equilibrium error. Notice that in equation (8), at time \(t\), the change in \(x_t, y_t\) will be driven in part by the short run dynamics (lagged \(\Delta x_t, \Delta y_t\)), and partly by the size and sign of the current disequilibrium, \(z_t\). Thus, the error correction model can be thought of as model which allows for disequilibrium, yet describes how that disequilibrium forces the economy back to equilibrium where \(x = ay\).

### 2 Estimation and Testing

From the perspective of economic theory, the forces of economic equilibrium should act as an attractor and force different variables to move together in the long run even though they may be individually nonstationary and hence vary widely in the short run. From a statistical point of view, the fact that we observe high correlations among economic time
series implies that the levels of macroeconomic time series tend to cluster around a linear attractor. Hence we should expect some pairs of macro time series to exhibit co-integration, such as consumption and income, money and income, yields on bonds of different maturity, spot and forward prices, etc.

2.1 Estimation of Co-integrated Models

Consider the problem of estimation of a bivariate system where we know the two series in question are $I(1)$ and $CI(1,1)$. The OLS co-integrating regression could be written as:

$$y_t = \alpha x_t + z_t.$$

(9)

2.1.1 Properties of CI Regressions

- The cointegrating parameter (bivariate case only) is unique.
- Superconsistency:

  We know that the OLS algorithm will minimize $V(\hat{z}_t)$. Thus, OLS will find a finite variance linear combination of $y, x$ if it exists. Because the variance of an $I(1)$ process tends to infinity as $t \to \infty$, and because only one linear combination can be stationary with a finite variance, least squares will find a good estimate of the co-integrating coefficient $\hat{\alpha}$. Because any linear combination other than the true co-integrating combination produces infinite residual variances, Stock (1987) showed that the least squares estimate will converge to its true value at the rate $T^{-1}$ as opposed to the usual (with stationary regressors) rate of convergence of $T^{-1/2}$.
The $R^2$ of a co-integrating regression will approach one asymptotically so the fit of the regression will essentially be perfect regardless as to whether $y$ is regressed on $x$ or vice versa. Although both series are treated as endogenous, co-integration ensures that in large samples, simultaneous equation bias is not a problem.

Super-consistent parameter estimates are obtained even if only a single equation from the bivariate system is estimated, and even in the presence of serially correlated errors.

- Although $\hat{\alpha}$ is superconsistent, Banerjee et al. (1993, p. 214-230) show that there can be substantial small sample biases and suggested estimating the long-run parameter using a dynamic regression as opposed to the static regression (9).

- Although $\hat{\alpha}$ is superconsistent, its asymptotic distribution depends on nuisance parameters arising from endogeneity of the regressor and serial correlation in the errors. As a result, normal procedures for conducting inference do not apply. For instance, t-ratios cannot be used to conduct inference, since they will not have a limiting normal distribution. Furthermore, standard methods of correcting for serial correlation—such as Cochrane-Orcutt corrections—will only make the estimator in the static model inconsistent. If it is absolutely necessary to conduct inference on the co-integrating parameters, (as opposed to focusing on the impact of the equilibrium error on the dynamics) then we must turn to the maximum likelihood method of Johansen (1990) or a host of other methods such as that of Phillips and Loretan (1991), or Phillips and Hansen (1990). A useful survey of estimation and testing issues in co-integrated systems is provided by Campbell and Perron ( “Pitfalls and Opportunities: What Macroeconomists Should Know About Unit Roots,” 1992 NBER Macroeconomic Annual).

If there is no correlation between shocks to $x_t$ and shocks to the cointegrating relationship, i.e. $x_t$ is truly weakly exogenous, and $\hat{z}_t \sim iid$, then the asymptotic distribution of $\hat{\alpha}$ is normal. (more on this later)
2.1.2 Dynamics

To estimate the more general dynamic model implied by a co-integrating relationship, we impose the constraint that the series co-integrate and write down an ECM such as in equation (8), repeated and expanded here for convenience.

\[
\Delta y_t = m_1 + \rho_1 (y_{t-1} - \hat{\alpha} x_{t-1}) + \sum_{l=1}^{k_1} \beta_{1,l} \Delta y_{t-l} + \sum_{l=1}^{k_2} \gamma_{1,l} \Delta x_{t-l} + \epsilon_{y_t} \tag{10}
\]

\[
\Delta x_t = m_2 + \rho_2 (y_{t-1} - \hat{\alpha} x_{t-1}) + \sum_{l=1}^{k_1} \beta_{2,l} \Delta y_{t-l} + \sum_{l=1}^{k_2} \gamma_{2,l} \Delta x_{t-l} + \epsilon_{x_t}
\]

If we know the value of \(\alpha\), it is straightforward to estimate the system defined in (10) by OLS (note SURE is the same as OLS on both equations because the same regressors occur in both equations). Engel and Granger (1987) suggested a two step procedure for estimating the ECMs defined above. They proved that the superconsistent estimate of \(\alpha\) obtained from the static co-integrating regression (9) can be used to form \(y_t - \hat{\alpha} x_t\) and then the equations in (10) estimated to obtain consistent estimates of the \(\beta'\)'s, \(\gamma'\)'s, and \(\rho'\)'s above. In fact, the superconsistency of the CI parameter \(\alpha\) implies that the parameter estimates in (10) are essentially the same as would be obtained if we used the true value of \(\alpha\). The Engle-Granger Two-step procedure is convenient because you do not have to specify the dynamics of the model until the error-correction structure is estimated. However, below we will discuss reasons that it may be preferable to begin with the dynamic specification at the outset.

2.1.3 Engel- Granger Two Step Procedure

1. Estimate the static co-integrating regression (9).
2. Take the residuals from that regression $\hat{z}_t$ and use them (lagged one period) in the ECM (10)

It is typical to include a constant in both the static CI regression as well as in the ECM. Also, with seasonal CI, we would have to include seasonal equilibrium errors, substantially complicating the analysis.

2.2 Testing for non-Co-integration

The procedure used for testing for co-integration is actually an extension of the ADF test to the case of estimated residuals. Thus the null hypothesis is that the estimated equilibrium error (the residuals from the static co-integrating regression) $\hat{z}_t = y_t - \hat{\alpha}x_t$ is a unit root process. If the series $x, y$ are not co-integrated, then there is no linear combination which is stationary. Hence, the residuals from the co-integrating regression will have a unit root in its AR polynomial. In contrast, if the series are co-integrated, the residuals will be stationary.

2.2.1 Step by Step Test for Co-integration

1. Estimate the CI regression, equation (9).
2. Conduct an ADF test on the residuals including an intercept (Note that if a constant is included in the cointegrating regression, but not in the ADF regression, you get the same results as a constant in the ADF reg but not the CI reg, etc.).
4. If the value of the test statistic is not sufficiently negative to reject the null of non-co-integration (a unit root in the equilibrium error), then estimate a model in differences. That is if you cannot reject non-co-integration, then you impose that restriction and set $\rho_1, \rho_2$ equal to zero in equation (10).
5. If the value of the test statistic is larger in absolute value than the critical value, then you reject the null of non-co-integration. It is common to then assert evidence of CI and estimate (10).

Notice that the critical values for testing the null of non-co-integration are now a function of the number of regressors included in the static co-integrating regression (9). This is because OLS tends to find increasingly white residuals as additional regressors are added to the CI regression.

There are at least three reasons the concept of co-integration is central to econometric modeling of macroeconomic series.

1. Co-integration formalizes a link between series with different orders of integration. That is, series of higher order of integration can be linked to lower order of integration series via a linear combination which reduces the order of integration of the former series. The linear combination is thought of as the equilibrium (statistically) relationship among the series.

2. Co-integrating relationships are identified as equilibrium relationships linking $I(d)$ series, whereas non-co-integrating regressions are spurious. In other words, regressions of levels of $I(d)$ variables make sense only if the variables are co-integrated. Thus, co-integration distinguishes meaningful from “nonsense” (Yule, 1926) or “spurious” (Granger and Newbold, 1974, 1977) regressions.

3. Finally, a set of co-integrated series will have an error-correction representation. In other words, the series can be modeled so that a term which represents the deviations of the observed values from a long-run equilibrium enters the model. The result provides a synthesis of the time-series method of differencing series to avoid spurious regressions, and the econometric methods which focus on equilibrium relationships in levels. Both of these methods could be seen as flawed: the former method discards all long run information in the levels of the series, and the later tends to ignore the spurious regression problem. ECMs provide a method of combining the advantages of working with stationary variables while not discarding levels information. (See Sargan (1964), Hendry and Anderson (1977), and Davidson et al. (1978).)
3 Estimation and Testing Hypotheses in CI Regressions

As described above, only in the very special case where cointegrating residuals are serially uncorrelated and there is no endogeneity problem will standard inference be possible for cointegrating parameters. Consider the following Triangular representation of the cointegrating system due to Phillips. Let $y_{1t}$ be a scalar, and $y_{2t}$ a possibly $(g + 1)$ vector such that,

$$y_{1t} = \alpha + \gamma' y_{2t} + z_t$$

(11)

$$y_{2t} = y_{2t-1} + u_{2t},$$

(12)

where both $y_{1t}$ and $y_{2t}$ are $I(1)$, but $z_t \sim I(0)$. In the special case where,

$$\begin{pmatrix} z_t \\ u_{2t} \end{pmatrix} \sim i.i.dN \left( \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma_1^2 & 0' \\ 0 & \Omega_{22} \end{bmatrix} \right),$$

(13)

so there is no serial correlation, and $E[z_t u_{2t}] = 0$, despite the $I(1)$ regressors, OLS estimates and standard t and F statistics are valid. This conclusion also extends to the case of serially correlated errors in (11) so long as there is still no cross equation error correlation, i.e. so long as $x_t$ is weakly exogenous. If $z_t$ is serially correlated, we know that the OLS estimate
of the variance is biased and needs to be replaced, for instance with a NW type estimator that allows for general serial correlation up to lag $q$,

$$
\hat{\sigma}_1^2 = \hat{c}_0 + 2 \cdot \sum_{j=1}^{q} [1 - j/(q + 1)] \hat{c}_j,
$$

(14)

for

$$
\hat{c}_j = T^{-1} \sum_{t=j+1}^{T} \hat{z}_t \hat{z}_{t-j},
$$

(15)

where $\hat{v}_t = y_{1t} - \hat{\alpha} - \hat{\gamma} x_t$ is the OLS residual from the estimation of the cointegrating regression. Replacing the usual estimate of the error variance with the corrected estimate from (14) will once again result in valid t and F statistics, so long as there is no cross-equation error correlation. Given the unlikely event that $x_t$ is truly exogenous, resulting in $E[\epsilon_t u_{2t}] \neq 0$, let $\epsilon_t = [z_t u_{2t}]'$, and

$$
E[\epsilon_t \epsilon_t'] = \begin{bmatrix}
\omega_{11} & \omega_{12}' \\
\omega_{21} & \Omega_{22}
\end{bmatrix},
$$

(16)

3.1 Saikkonen (91) and Phillips & Loretan (91)

Saikkonen (1991) suggested removing the effects of $\omega_{21} \neq 0$ by augmenting the cointegrating regression with leads and lags of $\Delta x_t$. Since $z_t$ and $u_{2t}$ are stationary, we know that it is possible to remove the effects of correlation between $z$ and $u$. For example, the residuals, $\tilde{z}$, from the following regression

$$
z_t = \sum_{s=-p}^{p} \beta_s' u_{2t-s} + \tilde{z}_t,
$$

(17)
are uncorrelated with \( u_{2t-s} \), for \( s = -p, -p + 1, \ldots, p \) by construction. Given the assumed triangular structure of the CI system, we can also remove these effects by augmenting the CI regression with leads and lags:

\[
y_{1t} = \alpha + \gamma' y_{2t} + \sum_{s=-p}^{p} \beta_s' \Delta x_{t-s} + \tilde{z}_t
\]  

(18)

To test a hypothesis about the value of the cointegrating vector \( \gamma \), you can estimate (18) by OLS, and calculate the usual t and F statistics. These statistics need to be calculated using NW std errors, so if your software does not allow for an automatic NW correction, you need to use equation (14) to replace the usual OLS esimtate for \( \hat{\sigma}^2 \).

3.2 FMOLS

Rather than estimating additional parameters on leads and lags as in (18), Phillips and Hansen (1990) use a non-parametric approach to eliminating the effects of endogeneity. They start with the Triangular system and assume that the error vector follows an infinite order MA process,

\[
\begin{pmatrix}
z_t \\
u_{2t}
\end{pmatrix} = \psi(L) \epsilon_t,
\]

(19)

where \( E(\epsilon_t \epsilon_t') = PP' \), or our matrix (13 ) above. To write the Long Run covariance matrix, define

\[
\Lambda \equiv \Psi(1) \cdot P
\]

(20)
Thus, $\Sigma$ is just the covariance matrix of the cumulative sum of the $\epsilon_t$, and is generally estimated using a NW estimator:

$$
\hat{\Sigma} = \hat{\Gamma}_0 + \sum_{v=1}^{q} / [1 - v/(q + 1)](\hat{\Gamma}_v + \hat{\Gamma}_v'),
$$

where

$$
\hat{\Gamma}_v = T^{-1} \sum_{t=v+1}^{T} \begin{bmatrix}
\hat{z}_t \hat{z}_{t-v} & \hat{z}_t \hat{u}_{2t-v}'
\hat{u}_{2t} \hat{z}_{t-v} & \hat{u}_{2t} \hat{u}_{2t-v}'
\end{bmatrix}
$$

and $\hat{z}_t$ is the residual from the cointegrating regression, while $\hat{u}_{2t} = \Delta x_t$.

To purge the CI regression of the effect of endogeneity, Phillips and Hansen form,

$$
y_t^\dagger \equiv y_t - \hat{\Sigma}_{21} (\hat{\Sigma}_{22})^{-1} \Delta x_t
$$

$$
z_t^\dagger \equiv z_t - \hat{\Sigma}_{21} (\hat{\Sigma}_{22})^{-1} \Delta x_t,
$$

and estimate the following equation by OLS,

$$
y_t^\dagger = \alpha + \gamma x_t + z_t^\dagger.
$$
The equation is estimated to obtain the residuals which are used in calculating a serial correlation correction used in forming the FMOLS estimator,

$$\hat{\gamma} = (x_t'x_t)^{-1}(x_t'y_t - T\delta^\dagger),$$

(28)

where $\delta^\dagger \equiv \sum_{v=0}^{\infty} E(u_{2t}z_{t+v}^\dagger)$.

4 General to Specific Modeling and the Advantages of Dynamic Models

In the 1970s, econometricians began to question the use of regressions in levels. Their worries were reinforced by the focus of time-series analysts and the seminal work of Box and Jenkins. Additional evidence of the need to carefully consider dynamics came from the failure of large scale econometric models (principally static levels regressions) to outperform single equation (dynamic) time series models in forecasting exercises. The General to Specific approach preferred models which began with as general a specification as possible and then simplified that model by imposing constraints that were consistent with the data observations. The findings of Engle and Granger returned some confidence in static levels regressions (at least when they are co-integrating). However, there are still a number of reasons to treat dynamic regressions as preferable to static ones.

---

2See Mizon (1977), Sims (1977), Hendry and Mizon (1978), and Hendry and Richard (1982) for examples.
1. Although the OLS estimate of the co-integrating vector is super-consistent, these estimators suffer from small sample bias that can be substantial. The bias is, at least in part, due to serial correlation in \((y_t, x_t)\) and the endogeneity of the two series. The bias is related to the mean lag of the dynamic adjustment of \(y_t\) to \(x_t\). Thus it is reasonable to think that estimating dynamic models is an appropriate way of dealing with small sample bias of static regressions.

2. Even if the cointegrating vector is consistently estimated, the distribution of the coefficient estimate will typically be non-standard. However, in dynamic models, standard asymptotic theory is often appropriate.

Single Equation Static regressions may suffer from at least one of the following five problems:

1. The non-standard distribution of coefficient estimates.

2. The error processes are likely not martingale difference sequences.

3. In combination with the previous problem, autocorrelation of the series \((y_t, x_t)\) results in small sample ‘second order’ biases.

4. When the relationship is between more than two series, there may be more than one cointegrating vector,

5. Weak exogeneity may fail if, for instance, a cointegrating vector belongs in more than one equation in the system.

By working with dynamic models, we hope to avoid the first three problems. To get around the last two, we will have to use systems methods. That is, if weak exogeneity fails, the processes generating the series \(x_t\) must be estimated.
4.1 Balance Again

As discussed above, a requirement of estimation with integrated variables is balance in the orders of integration of the variables on the left-hand side and right-hand sides of the regression equation. Even when our equations balance, test statistics from such regressions will typically have non-standard distributions. Thus, ordinary tables of Critical values are often useless.

In contrast, if the equation balances by having both sides $I(0)$, then conventional testing methods are appropriate. One way to balance an equation is to look for a co-integrated set of regressors and use a sufficiently differenced regressand. Because lagged variables used as regressors often have the same effect as providing a co-integrated set of regressors, there is a particular advantage to beginning with general dynamic specifications. The goal is to find some way to re-specify the equations so that the regressors are either jointly or individually integrated of order zero, while the dependent variable is $I(0)$ as well. The important point for inference is that it be possible to perform the transformation.\(^3\) The possibility of such transformations obviously is enhanced by following a general to specific approach to model building. The GTS method is effective for a very simple reason: “the inclusion of several variables and their lags as regressors increases the chances of obtaining a co-integrated set of regressors. A dependent variable made stationary by differencing can be regressed on this co-integrated set, and standard $t$-, $F$-, and normal tables can be used for inference. The regression would then take the form of a differenced variable as the regressand, and differences and levels of the variables as regressors.” (Banerjee et al. (1993, p. 168))

\(^3\)That is, the transformation does not actually have to be carried out. It is essential only that the transformation is possible because the appropriate regressors are present. Regressions that are linear transformations of each other will have identical statistical properties.
The general rule of thumb for inference in balanced dynamic regressions is that estimators of parameters which can be rewritten as coefficients on mean-zero, stationary regressors, will have asymptotically normal joint distributions. The re-specification to stationary form can be accomplished either by differencing, or by linear combination of co-integrated regressors. These conclusions follow from the work of Sims et al. (1990). Here we will first outline the main points of Sims et al. found in Banerjee and Dolado (1988). Next we look briefly at the “Tale of Two Econometricians” from Stock and Watson (1988). Consider the following regression:

\[
\Delta c_t = \sum_{i=1}^{n-1} \alpha_i \Delta y_{t-i} + \sum_{i=1}^{m-1} \gamma_i \Delta c_{t-i} + \beta y_{t-n} + \delta c_{t-m} + \mu_t,
\]

where \( y_t \) is the log of disposable income and \( c_t \) is the log of consumption. Both series are assumed to be \( I(1) \) in levels. Thus, although we have \( I(1) \) regressors, if they are co-integrated (as they must be if the PIH holds), then equation (29) is \( I(0) \) balanced. This stationary balance ensures that t-tests on the coefficients of all regressors are asymptotically normal. While the individual t-tests are normally distributed, joint tests on the coefficients on \( I(1) \) variables have non-standard distributions. The main conclusions that can be proven from Theorems 1 and 2 in Sims, Stock, and Watson (1990) are:

1. The t-statistic for each individual coefficient is asymptotically normally distributed.

2. The F-statistics for testing the joint significance of any subset of the stationary regressors have standard asymptotic distributions. Thus, we can test the joint significance of any subset of \( \Delta y_{t-i} (i = 1, \ldots n-1) \) and \( \Delta c_{t-i} (i = 1, \ldots m-1) \) using standard F-tables.

3. If the regressors in the set of \( I(1) \) variables are co-integrated and are combined to give \( p \) stationary regressors and \( q \) nonstationary regressors, then an F-statistic for testing the

\footnote{For examples see Sims et al. (1990, p. 119), Banerjee and Dolado (1988), Stock and West (1988), and Stock and Watson (1988).}
joint significance of any of the $p$ derived (linear combinations of co-integrated variables) stationary regressors along with any of the original stationary regressors will also have a standard asymptotic distribution.

4. F-statistics for the joint significance of any subset of the I(1) regressors in combination with any of the I(0) regressors will have a non-standard distribution.

The first point obtains due to the co-integration of the $I(1)$ regressors. In other words, the equation is balanced and stationary. Also, it is necessary that both $\delta$ and $\beta$ can be written as coefficients on stationary regressors as shown below. The second and third points are straightforward since they simply involve joint tests on stationary variables in a balanced regression. Finally, the fourth point is important because it contrasts with the standard distribution for the individual t-statistics.

Consider the simplified version of equation (29):

$$\Delta c_t = \delta c_{t-1} + \beta y_{t-1} + \mu_t, \quad (30)$$

Now add and subtract $\beta c_{t-1}$ from the LHS of equation (30) and rearrange:

$$\Delta c_t = \gamma_1 c_{t-1} + \gamma_2 s_{t-1} + \mu_t, \quad (31)$$

where $s_{t-1} = y_{t-1} - c_{t-1}$, $\gamma_1 = \beta + \delta$, $\gamma_2 = \beta$. Note that if the PIH is true, then $s_{t-1} \sim I(0)$. In the form (31), the t-statistic for $\gamma_1$ is functional of Wiener processes (i.e., non-standard), while $t_{\gamma_2}$ is asymptotically normally distributed.
4.1.1 A Tale of Two Econometricians —Stock and Watson (1988)

Stock and Watson begin with a simple DGP based on the PIH:

\[ y_t = y_t^* + y_t^s \] (32)
\[ y_t^* = y_{t-1}^s + \mu_t \] (33)
\[ c_t = y_t^s \] (34)
\[ p_t = p_{t-1} + v_t \] (35)

where \( E[u_tv_t] = 0 \), and

- \( y_t^* \) = the random walk (unobserved) permanent component of disposable income.
- \( c_t \) = consumption
- \( y_t^s \) = transitory (stationary) component of disposable income.
- \( p_t \) = price level assumed to follow a random walk.

4.1.2 Misguided Econometrician

The misguided econometrician (ME) who ignores the issues associated with nonstationary processes estimates the following regressions:

\[ c_t = \alpha_1 + \beta_1 p_t + \epsilon_{1t} \] (test money illusion) (36)
\[ c_t = \alpha_2 + \beta_2 t + \epsilon_{2t} \] (test for trend in consumption) (37)
\[ \Delta c_t = \alpha_3 + \beta_3 \Delta y_t + \epsilon_{3t} \] (calculate MPC) (38)
Each of these regressions are invalid. The first regression is spurious since consumption and price are not cointegrated. Thus, when the ME finds an extremely high $R^2$ and large $t$-statistic, it is meaningless. The ME thought he was testing for homogeneity, while rejecting the (true or false) null is not a reason for accepting the alternative of money illusion. The second regression (estimated to test for a trend in consumption) is also spurious since consumption follows a random walk and hence $\epsilon_{2t}$ is a detrended random walk. The third regression involves first differencing of integrated and co-integrated variables and thus is mis-specified. Furthermore, as an attempt to estimate the marginal propensity to consume, it generates downward biased estimates. Note that $\Delta y_t = \Delta y^*_t + \Delta y^*_s$. Finally, the fourth regression is unbalanced since it tries to explain an $I(0)$ process with a series that contains a stochastic trend. Such a regression results in trivial cointegration —such that the only coefficient possible for $\beta_4$ is zero, otherwise the error would contain a unit root. However, the $t$-statistic in the fourth equation is biased away from zero and may lead to non-rejection of the null of insignificance (thus rejection the PIH).

4.1.3 Savvy Econometrician

In contrast to the Misguided Econometrician, the savvy econometrician (SE) estimates the following regressions:

\[
\Delta c_t = \alpha_4 + \beta_4 y_{t-1} + \epsilon_{4t} \quad \text{(test PIH)}
\]  

(39)

\[
\Delta c_t = \gamma_1 + \delta_1 y_t + \epsilon_{1t}
\]

(40)

\[5\text{Nelson and Kang(1981) demonstrated that attempting to render I(1) series stationary by detrending results in spurious relationships.}\]
\[ c_t = \gamma_2 + \delta_2 c_{t-1} + \delta_3 c_{t-2} + \epsilon_{2t} \]  
\[ c_t = \gamma_3 + \delta_4 c_{t-1} + \delta_5 y_{t-1} + \epsilon_{3t} \]  

Each of these regressions generally allow for valid inference. In the first regression, co-integration holds, so that the coefficients are super-consistent, but will have a Wiener distribution and standard errors are generally biased due to autocorrelated residuals. The second regression can be re-parameterized as \( c_t = \gamma_2 + (\delta_2 + \delta_3) c_{t-1} - \delta_3 (c_{t-1} - c_{t-2}) + \epsilon_t \). Therefore, the equation can be written such that \( \delta_3 \) is the coefficient on a stationary variable. The same is true of \( \delta_2 \) (in a different re-parameterization). Thus, t-statistics on individual coefficients have standard distributions. This argument also holds for the third equation, but this time the stationary re-parameterization involves the co-integrating relation \( y_{t-1} - c_{t-2} \).

The lesson to be learned from this story is the necessity of keeping track of the orders of integration of both sides of your regression equations, and follow a GTS approach. A model with an overly restrictive dynamic structure is likely to produce misleading inference simply because of the inconsistent orders of integration. The ME could have overcome many of his difficulties by beginning with a more general specification such as

\[ c_t = \beta_0 + \beta_1 t + \beta_2 p_t + \beta_3 y_t + \beta_4 c_{t-1} + \beta_5 y_{t-1} + \beta_6 p_{t-1} + \epsilon_t \]  

4.2 Advantages of Dynamic CI tests

In the preceding section, we saw that an econometrician beginning with an excessively restricted model is more likely to reach incorrect conclusions due to invalid inference on coefficients of interest. It turns out that the same basic conclusion holds when testing for
cointegration. Based on simulation evidence presented in Banerjee et al. (1993), the following conclusions emphasize the advantages of dynamic specifications:

1. Dynamic regressions tend to produce lower biases in estimates of the CI parameter. This conclusion holds even when the DGP is a simple one, such that the dynamic model is substantially overparameterized.

2. The largest biases exist for unbalanced and nearly unbalanced regressions. Of course, unbalanced regressions are likely to have asymptotically divergent coefficient estimates and t-statistics (see Phillips (1986)).

3. Underparameterized dynamic regressions perform substantially worse than both correctly-parameterized and over-parameterized models. This suggests the use of rich dynamic structures.

In summary, although OLS estimators in static co-integrating regressions are super-consistent, substantial bias exists even in moderate samples. Specifying dynamic regressions helps to alleviate this problem. The usefulness of dynamic specifications is reinforced by their ability to detect co-integration even when standard EG two step procedures fail. The basic rational is that the use of ADF tests on residuals from static regressions imposes a common factor restriction where it may not exist. Working with a general dynamic specification and testing for cointegration imposes no potentially invalid restrictions. In fact, one way of examining whether cointegration holds is to test for common factors with a root of unity. In other words, test whether the AR polynomial in the residuals has a unit root shared by the AR polynomials of the dependent and independent variables in your model. If this hypothesis cannot be rejected, then the data must be first differenced to remove the unit root in the residuals. There are a number of methods of testing for cointegration in dynamic specifications. One is to write a general ECM and test whether the parameter on the equilibrium error is significantly different from zero. If not, then the ECM is a growth rate model with no equilibrium feedback.
To see the advantage of testing for co-integration in a dynamic framework, consider the
simple ADL(1,1) model we first considered in LectureNotes1 (Section 4.3, equation (30)).

\[
y_t = \alpha_1 y_{t-1} + \beta_0 z_t + \beta_1 z_{t-1} + \epsilon_t, \quad \epsilon_t \sim IN[0, \sigma^2]
\]  

(44)

Now suppose that instead of estimating equation (44), we follow the EG two step procedure
by first estimating the static model

\[
y_t = K z_t + u_t,
\]

and then test the null of no co-integration by DF test using \( \hat{u}_t = y_t - \hat{K} z_t \). Thus, the DF
regression can be written as:

\[
\Delta(y_t - \hat{K} z_t) = \rho(y_{t-1} - \hat{K} z_{t-1}) + w_t
\]

(46)

The DF based test for no co-integration imposes a common factor on the dynamic structure
of the ADL(1,1) in equation (44). To see this, recall that the common factor restriction
is given by \( \beta_0 \alpha_1 + \beta_1 = 0 \), and consider the following isomorphic re-parameterization of
equation (44).

\[
y_t = \beta_0 z_t + \alpha_1(y_{t-1} - \beta_0 z_{t-1}) + (\beta_0 \alpha_1 + \beta_1) z_{t-1} + \epsilon_t
\]

now impose the COMFAC restriction to get:

\[
y_t - y_{t-1} = \beta_0 z_t - \beta_0 z_{t-1} - y_{t-1} + \beta_0 z_{t-1} + \alpha_1(y_{t-1} - \beta_0 z_{t-1}) + v_t
\]

\[
\Delta y_t = \beta_0 \Delta z_t + (\alpha_1 - 1)(y_{t-1} - \beta_0 z_{t-1}) + v_t
\]

(47)
Finally, notice that adding $\hat{K}\Delta z_t$ to both sides of the DF regression in equation (46), and letting $\rho = \alpha - 1$, produces (47). In other words, a common factor model with the restriction that the long run and short run adjustment are identical, $K = \beta_0$. This means that if any adjustment towards equilibrium occurs, it will be instantaneous. Because the comfac restriction implied by the DF regression may be invalid, a number of authors have suggested testing for the null of no co-integration in a model with more richly specified dynamics.

Recall that regardless of whether you estimate a static or a dynamic regression to examine cointegration, the regressors need to be strongly exogenous. So in this example (due to BDGH, 1998), we consider a simple DGP where the $z_t$ variable is strongly exogenous for the regression parameters by construction.

\begin{align*}
y_t &= \alpha_1 y_{t-1} + \beta_0 z_t + \beta_1 z_{t-1} + \epsilon_t \quad (48) \\
z_t &= z_{t-1} + \nu_t \quad (49)
\end{align*}

Where we assume that $y_t$ and $z_t$ are CI(1,1), and economic theory tells us that $\alpha_1 + \beta_0 + \beta_1 = 1$, in other words, we will assume that $K = \frac{\beta_0 + \beta_1}{1 - \alpha_1} = 1$. Rewrite equation (48) by subtracting $y_{t-1}$ from both sides of the equality, then add and subtract $\beta_0 z_{t-1}$ from the right side to get:

\begin{align*}
\Delta y_t &= (\alpha_1 - 1) y_{t-1} + \beta_0 \Delta z_t + (\beta_0 + \beta_1) z_{t-1} + \epsilon_t \\
&= (\alpha_1 - 1) (y_{t-1} - z_{t-1}) + \beta_0 \Delta z_t + \epsilon_t \quad (50)
\end{align*}

using our theory assumption: $1 - \alpha_1 = \beta_0 + \beta_1$

\begin{align*}
\Delta y_t &= (\alpha_1 - 1) (y_{t-1} - z_{t-1}) + \beta_0 \Delta z_t + \epsilon_t \quad (51)
\end{align*}
By our CI(1,1) assumption (and the theory restriction), we know $y_t - z_t \sim I(0)$ as is the differenced $z_t$ regressor. So we have an error correction model with our equilibrium error given by $y_t - z_t$. Now, consider a slightly different isomorphic version of equation (48).

First add $y_{t-1}$ to both sides and then add and subtract $\Delta z_t$ to the right hand side of (48).

$$y_t = \alpha_1(y_{t-1} - z_{t-1}) + z_{t-1} + \beta_0 \Delta z_t + \epsilon_t$$

$$y_t = \alpha_1(y_{t-1} - z_{t-1}) + z_{t-1} + (z_t - z_{t-1}) - \Delta z_t + \beta_0 \Delta z_t + \epsilon_t$$

$$y_t = z_t + \alpha_1(y_{t-1} - z_{t-1}) + (\beta_0 - 1) \Delta + \epsilon_t$$

Where $K = 1$, the parameter on $z_t$ is set equal to unity due to the theory restriction imposed above. Notice that estimation of the static model in (45) forces the residuals to capture all of dynamics of the equilibrium error, $\alpha_1(y_{t-1} - z_{t-1})$ and the dynamics of $(\beta_0 - 1) \Delta z_t$. So $u_t = \alpha_1(y_{t-1} - z_{t-1}) + \beta_0 \Delta z_t + \epsilon_t$ will be highly serially correlated. The bias in the static regression estimate of $K$ can be shown to be a function of the Long Run variance of the residuals from the static regression, $\sigma^2$. This long run variance in turn may be written as

$$\sigma^2 = \sigma_u^2 + 2\lambda,$$

where $\sigma_u^2 = E[u_t^2]$, and $\lambda = \sum_{i=0, i \neq j}^{\infty} E[u_i u_j]$. In this case where we have no “endogeneity problem” by construction, the bias in the static parameter estimate arises from the serial correlation in the static error term, i.e. $\lambda \neq 0$. A simple way to reduce the bias is to reparameterize the static equation so that $\lambda$ is set to zero. In otherwords, actually estimate
the terms $\alpha_1(y_{t-1} - z_{t-1}) + (\beta_0 - 1)\Delta z_t$ rather than leave them to the residuals. Generally, equation (52) is estimated

$$\Delta y_t = \gamma_1 \Delta z_t + \gamma_2 (y_{t-1} - z_{t-1}) + \gamma_3 z_{t-1} + w_t$$  \hspace{1cm} (52)

where $\gamma_1 = \beta_0$, $\gamma_2 = \alpha_1 - 1$, and $\gamma_3 = \alpha_1 + \beta_0 + \beta_1 - 1$. This last parameter on the lagged level of the strongly exogenous regressor serves to break the homogeneity assumption in the “equilibrium error” $(y - z)$. Note that (52) is isomorphic to an error correction representation, where $K$ is not restricted to unity, and is used to form the equilibrium error $y_t - K z_t$. An estimate of the cointegrating parameter can be obtained from estimation of (52) as $K = 1 - \gamma_3 / \gamma_2 = \frac{\alpha_1 + \beta_0 + \beta_1 - 1}{\alpha_1 - 1}$, and $\gamma_2$ still provides a measure of the speed of adjustment.

In fact, the PcGive co-integration test is exactly such a test conducted on the general ADL(p,q). This test augments the ECM with a redundant regressor which avoids the need to know the true value the co-integrating vector $\hat{K}$ (See Park and Phillips, 1988, Banerjee and Hendry, 1992, and Banerjee et al., 1993).

The co-integration test is a pseudo t-test of the null hypothesis of no co-integration, $H_0 : \gamma_2 = 0$. Under the null of no co-integration, the t-test for $\gamma_2 = 0$ is a functional of Wiener distributions, and the exact limiting distributions are derived in Kiviet and Phillips (1992). PcGive calculates the appropriate critical values and reports significance levels in the PcGive Unit Root tests (see Hendry and Doornik (1996, p 235)).

Banerjee et al, (1993) demonstrate that when a common-factor restriction is close to valid, the ADF test for no co-integration does relatively well. However, when the comfac approximation is poor, the ADF test regression has a large increase in its residual standard error and the test does poorly. Finally, it is important to note that the dynamic co-
integration test relies on the strong exogeneity of \( z_t \), i.e. there must be no feedback from \( y_t \) onto current \( z_t \), and the ECM must enter only the equation for \( y_t \), not the equation for \( z_t \). If \( z_t \) is not at least weakly exogenous, the test will do poorly.

4.3 Model Validity Tests—part of a GTS modeling approach

An important part of the GTS methodology is the selection of a tentatively adequate data generating process. This requires that our general specification be congruent with the available evidence, and in LectureNotes1 I outlined the necessary (but not sufficient) conditions for a model to be deemed congruent. I will repeat those conditions below, but first, it is important to reiterate that statistical analysis follows an ‘as if’ strategy of model selection. In other words, we treat the existing model ‘as if it were the true DGP’ and then apply various test statistics which have known distributions when the null hypothesis is valid. If the test result is consistent with its distribution under the null, then the model is *temporarily retained* —if not, the model is *revised*. Because a rejection of a test does not have unique implications, the process of revising models is often painfully slow, and subject to luck and uncertainty. Divine inspiration or guesses may be required! It is not uncommon for economists (particularly theorists who have never tested their theories against real world data) to refer to such a process, pejoratively as ‘data-mining’. Yet, “from the fact that it is of no scientific or logical significance how a theory is arrived at it follows that no way is illegitimate” (Magee 1982, p. 32).

Finally, if a model has been selected through data-based revision to satisfy all necessary criteria, it can be *genuinely* tested only on (*genuinely*) new data or against new criteria or new rival models. The values of test statistics which we often report along with parameter estimates and hypothesis tests serve mainly to illustrate the validity of the model selection
process as opposed to the model itself. Until we have been able to test our models against new data evidence, it is hazardous to place undo value on their implications — no matter how appealing these implications may seem.

Hendry and Richard (1983) suggested six criteria that an acceptable model should meet. 6

1. Models must be data admissible.
2. Models must be consistent with theory.
3. A satisfactory model must be data coherent.
4. Satisfactory models should have regressors that are (at least) weakly exogenous.
5. Satisfactory models must exhibit parameter constancy.
6. A satisfactory model should encompass a wide range of competing models.

4.3.1 Data Admissible

For example, suppose you are trying to study the effect of raising the A-plus fees on the proportion of low income families whose children attend A-plus programs. The ratio of participating to eligible (low income) children must lie between 0 and 1. Thus, if the dependent variable in your model can lie outside that range, then your model would be capable of predicting that less than zero percent (or more than 100 percent) of the eligible low income children attended A-plus. Examining a coefficient on the A-plus fee in such a regression would not be very useful.

6See also Charemza and Deadman (1997), Gilbert (1986), Doornik and Hendry (1996), and Hendry (1993, ch. 18)
4.3.2 Theory Consistent

This is relatively straightforward. The model must be consistent with at least one theory model. Given the proliferation of theoretical models, this does not seem overly demanding. (On the other hand...!)

4.3.3 Data Coherent

Two separate aspects of data coherency are the closeness of the description of the data, and the lack of any systematic deviations from the data. In other words, not only should the model fit the data closely, but the difference between the fitted values from the model and the actual data must be random in the sense that they should not be predictable based on their own past. If this is not true, then it is straightforward to predict how a model will misforecast, and therefore to improve the model by simply augmenting the original model with the model for explaining the misforecasts.

A typical method of evaluating goodness of fit is some scalar measure such as $\bar{R}^2$, an information criteria, etc. A useful descriptive statistic is the residual standard deviation as a percentage of the level of the dependent variable. A nice feature of such measure is its correspondence with the principle of ‘variance dominance.’ Theil (1971, p 543) has shown that a ‘true model’ will variance dominate ‘false models’.

Serial correlation of a model’s residuals is a departure from randomness, and thus models with serially correlated errors are not data coherent. Thus, a necessary condition for a non-systematic residuals is that they be white noise. White noise processes cannot be predicted linearly from their own past alone. Testing for white noise is done using a Lagrange Multiplier approach by estimating an auxiliary regression. PcGive provides tests for white noise both
4.3.4 White Noise and Structure

It is important to realize that white noise residuals are not necessarily unpredictable. White noise may have structure (e.g., a model is likely to underpredict peaks and overpredict troughs) such as \( u_t = \epsilon_{1t} + \gamma \epsilon_{2t-1} \), where the \( \epsilon_{it} \) are unrelated white noise processes. Therefore, white noise may be predictable from information that is not included in the model, such as \( \epsilon_{2t-1} \) (see Granger(1983)). Omitting a variable that is nearly white noise will not lead to a rejection of the null of serially independent residuals, but the residuals will still be predictable with the correct information set, and the model may exhibit parameter instability if the marginalized white noise process changes. Furthermore, residuals which appear white may even be predictable based on functions of lagged right hand side variables. This will occur when an invalid common factor restriction has been imposed via a Cochrane-Orcutt correction.

4.3.5 Innovations

A stronger condition than white noise is that residuals be an innovation, i.e. unpredictable both from their own past and from the past of other variables.

"Definition1 \( \epsilon_t \) is a (mean) white noise process if \( E[\epsilon_t|E_{t-1}] = E[\epsilon_t], \forall t \), where \( E_{t-1} = (\epsilon_{t-1}, \epsilon_{t-2}, \ldots \epsilon_0) \). Thus, \( \epsilon_t \) is not predictable linearly from its past beyond its unconditional mean (taken to be zero)
Definition 2

(i) \( v_t \) is a mean innovation process (MIP) with respect to an information set \( I_t \) if \( E[v_t | I_t] = 0, \forall t; \) and (ii) \( v_t \) is an innovation process with respect to \( I_t \) if it is a MIP and \( D(v_t | I_t) = D(v_t), \forall t. \" \) (See Hendry and Richard (1982)).

Whether or not \( \epsilon_t \) is an innovation process depends on the information set considered. White noise is a MIP relative to its own past, yet if we add other variables to the information set, it may no longer be a MIP. It will be only if the new variables do not Granger-cause \( \epsilon_t \).

We start with a general model at the outset, and take it’s residual variance as a measure of the ‘innovation variance’. To test whether a restricted model’s errors are innovation processes, we simply test that the restricted model does not fit significantly worse than the general specification. In other words, we test that the restricted model is not variance dominated by the unrestricted model. Typically we use F-tests to test the exclusion restrictions that generate the restricted model. See Doornik and Hendry (1996, p 241 (hereafter DH).)

4.3.6 Heteroscedasticity

If you observe a large increase in your residual standard deviation, a possible cause is invalid marginalization. Other causes include measurement error in the data (particularly if it changes over time), or more variation in the real world. A noisier world — or changing data accuracy may show up in the form of residual heteroscedasticity. PcGive provides three tests for heteroscedasticity and also allows for heteroscedastic consistent standard errors (see Doornik and Hendry 1996, p. 240.) Finally, a simple indicator of residual outliers is to calculate skewness and excess kurtosis of the residuals (see DH p. 209)
4.3.7 Weakly Exogenous Regressors

As discussed above, the weak exogeneity criteria implies that conditioning on the regressor set is valid. In contrast, if the regressors are not weakly exogenous, then they are endogenous and must be modeled jointly. It does not make sense to claim you have ‘explained’ $y$ in terms of $z$, if $z$ is also ‘explained’ in terms of $y$.

Note that weak exogeneity is only an adequacy criteria for hypothesis testing and is not sufficient for forecasting or policy analysis. To see this, consider the following simple model,

\begin{align*}
y_t &= \alpha + \beta z_t + \mu_t, \quad \mu_t \sim NID(0, \sigma^2_{\mu}), \\
z_t &= \gamma + \delta y_{t-1} + \epsilon_t, \quad \epsilon \sim NID(0, \sigma^2_{\epsilon})
\end{align*}

(55)

(56)

where $E(\mu_t \epsilon_t) = 0 \forall s, t$. In this simple model, $z_t$ is weakly exogenous relative to the parameter $\beta$ which can be consistently estimated by using OLS on equation (55). Note, variables are not weakly exogenous with respect to each other, but rather with respect to the parameters of interest. However, if we want to use (55) to forecast more than a single period into the future, we also need to know the submodel (56) because any movement in our forecast in $y_t$ will feedback onto $z_{t+1}$. Thus, while weak exogeneity is generally all that is need to ensure consistent estimation of $\beta$, for forecasting we will need strong exogeneity.

4.3.7a Strong Exogeneity is the combination of weak exogeneity and the absence of feedback. Recall that in our simplification of the DGP in Seminar 1, we factored the DGP into three components. We wrote the PDF for the $z$ variables of interest as depending on both lagged $y$ and $z$, $D_3(z_t|Y_{t-1}, Z_{t-1}; \theta_3)$. However, for the $z$ variables to be strongly exogenous
for the parameters of interest $\theta_2$, we would have to be able to write $D_3$ as $D_3(z_t, Z_{t-1}; \theta_3)$. In other words, the $z$ variables do not depend on any past or present values of $y$. This condition is known as Granger non-causality. Thus, strong exogeneity is the combination of weak exogeneity plus Granger non-causality, see Engle, Hendry, and Richard (1983).

4.3.7b Super exogeneity is even stronger. It requires strong exogeneity plus the condition that the parameters of interest are independent of the process that generates the variables of interest! To see this intuitively, look again at (55). Suppose that $z_t$ is a vector of variables, some of which are expectations variables —such as consumers expectations of future taxes. This expectation might be modeled by including past income tax collections in the model. Now, suppose that these past personal tax collections are strongly exogenous for $\beta$. (This may not be true, but bear with me) Given strong exogeneity, (55) may forecast very well, but using it to study the effects of a change in tax policy would be misleading. If the tax authority changes its tax rates or the way it collects its taxes, then consumers are likely to respond by changing their behavior in response to the same level of taxation. In other words, if the process generating $z$ changes, then $\beta$ may very well change too. For such models to be useful for policy analysis, they would need to be re-specified in terms of more fundamental super exogenous variables.

Testing

Testing for exogeneity in a SEM is difficult —yet an indirect analysis is possible on the basis of parameter constancy tests. To fully test for exogeneity requires estimating the marginal process for $z_t$ and will be addressed again when we consider systems in Seminars 4 and 5.
Data Coherence and Weak Exogeneity deal with the issues of past and contemporaneous information respectively. Having dealt with those issues, it is also important to consider ‘relative future’ information. To see the importance of future information, consider rewriting our simple ADL model in matrix form:

\[ y = Z\beta + \epsilon. \] (57)

Now, partition the data set into two subsamples \((y_1Z_1)\) and \((y_2Z_2)\), and suppose that the following conditions hold:

\[
\begin{align*}
T_1^{-1}Z_1'Z_1 & \neq T_2^{-1}Z_2'Z_2 \\
T_1^{-1}Z_1'y_1 & \neq T_2^{-1}Z_2'y_2
\end{align*}
\] (58)

yet \(\hat{\beta}_1 \approx \hat{\beta}_2\). The expressions in equations (59) are intended to convey substantial differences in subperiod data correlations. Thus, although there have been substantial changes in the data correlations, \(y\) remains \(Z\beta\) plus an orthogonal disturbance. Finally, notice that a finding of parameter constancy is a necessary but not sufficient condition for model validity since it is possible that sub-period data movements are very similar.
The numerous methods of testing parameter constancy using recursive residuals to construct various chow type test were discussed in Seminar 2. One important point to note is that the $\chi^2$ form of Forecast test will always be larger than the F-form of the Chow test. In fact, the $\chi^2$ forecast test has some drawbacks as statistical test (see Kiviet, 1986), and therefore should simply be used as an index of the model’s usefulness for ex-ante forecasting. If the $\chi^2$ test is large, while the F-test does not reject constancy, then forecasts would likely be inaccurate even though the underlying (imprecisely estimated) parameters did not change. Finally, a useful (although not often available) means of testing parameter constancy is to fit the model to a pre-sample data set!

4.3.9 Encompass rival models

The proliferation of competing explanations (models) of empirical phenomena leads to important methodological issues. For a researcher to claim that their model, model $M_1$, is congruent, it must be a valid reduction of the DGP. In fact, empirical models are simply reduced re-combinations of the data. The question is which reduction/re-combination is the best approximation to the DGP. If model $M_1$ is supposed to explain the data, then it should also be able to explain the re-combination of the data represented by the rival model $M_2$. Furthermore, if there are two or more competing explanations for a given phenomenon, at least one of them must be wrong, and possibly all are.

Encompassing attempts to resolve the problem of a proliferation of competing models by requiring that any chosen model be capable of explaining the results of its rivals.

There are two simple reasons for augmenting the traditional testing strategy with encompassing tests:
1. rejection is not final
2. corroboration is not definite

When we test our models and reject them, it is nearly always possible to rescue the model by patching it in some way to eliminate the rejection. For example, using an autocorrelation correction to ensure our residuals are white noise. In fact, empirical modeling exercises typically continue so long as no rejection occurs. In that case, theories can never be rejected, since the rejections are not reported! Only the non-rejected model is reported. And the only way to reject the model is with new data evidence.

At the same time, corroboration is not definitive. A theory or hypothesis is corroborated by evidence that is consistent with the hypothesis when it could have rejected the hypothesis. But corroboration is well known provide only weak support and to be insufficient to verify an hypothesis. There are numerous corroboration paradoxes. For example, the proposition that “all As are Bs” implies that “all not Bs are not As”. So any piece of evidence (such as C through Z) where a “not B” is also “not A” corroborates the (false) proposition. Another problem is that researchers tend to think of the accumulation of corroboration evidence as providing a strengthening of support for a hypothesis. So the more examples of empirical work which find the LR MPC is equal to unity, the more evidence we have of the PIH. Yet multiple pieces of corroborating evidence may actually reject an hypothesis. For example, suppose that we have a simple theory that expects $\alpha + \beta < 0$, and $\alpha > 0$. Now suppose that research # 1 generates evidence that $\alpha = 2$. This evidence is clearly consistent with the theory and corroborates it. Now researcher # 2 produces evidence that $\beta = -1$. Individually, this piece of evidence is consistent with the theory, but together, these two corroborating pieces of evidence reject the theory.
The problem is that corroboration and non-rejection are necessary but insufficient. If they are not satisfied, the hypothesis is rejected or the theory is inadequate, but if satisfying them does not imply a true hypothesis. A strategy of postulating models and either corroborating them or rejecting them is an inadequate methodology.

Encompassing is intended to address the problem of a proliferation of competing models. It does this by requiring that any chosen model should be capable of explaining the results of its competitors. Encompassing attempts to improve the research strategy by building on earlier results to try and increase our ability to explain economic phenomena.

**Encompassing and Mis-Specification analysis**

Starting from knowledge of the DGP, the mis-specification of a false model can be predicted from knowledge of the DGP. Suppose the DGP is model $M_0$, and the false model is $M_1$.

\[
M_0 : \ y_t = \beta_1 x_{1t} + \beta_2 x_{2t} + \epsilon_t \quad (59)
\]

\[
M_1 : \ y_t = \alpha x_{1t} + v_t \quad (60)
\]

Rewrite the DGP as:

\[
M'_0 : \ y_t = (\beta_1 + \beta_2 b_{21}) x_{1t} + \beta_2 (x_{2t} - b_{21} x_{1t}) + \epsilon_t \quad (61)
\]

\[
= \alpha x_{1t} + v_t \quad (62)
\]
where \( b_{21} = m_{21}m_{11}^{-1} \), and \( m_{ij} = E[x_{it}x_{jt}] \). In other words, if we form the “completing model”

\[
x_{2t} = b_{21}x_{1t} + u_t, \tag{63}
\]

\( b_{21} \) is the parameter estimate from regressing \( x_{2t} \) on \( x_{1t} \), and \( u_t = (x_{2t} - b_{21}x_{1t}) \perp x_{1t}, \) so that \( v_t \perp x_{1t} \) in (62). Marginalizing with respect to \( x_{2t} \) in the \( M_0 \) is equivalent to imposing \( \beta_2 = 0 \) and therefore eliminating \( (x_{2t} - b_{21}x_{1t}) \) to generate \( M_1 \). Because we know the DGP, we also know that in model \( M_1 \),

\[
\alpha = \beta_1 + \beta_2 b_{21}. \tag{64}
\]

So knowledge of one model which we claim represents the DGP (or in this case knowledge of the DGP) might allow us to explain why \( \alpha \) has the “wrong” sign compared with that expected (known) for \( \beta_1 \). For example, if \( \beta_2 < 0 \), and \( b_{21} > 0 \) and sufficiently large to give \( |\beta_2 b_{21}| > \beta_1 \), then \( \alpha < 0 \), even though \( \beta_1 > 0 \).

A model, \( M_1 \), that is claimed to be congruent, must encompass (explain the parameters of) other models. When it does, the other models are redundant. Consider a second model, \( M_2 \).

\[
M_2 : \ y_t = \delta x_{2t} + u_t \tag{65}
\]

from \( M_1 \) we know

\[
y_t = \beta_1(x_{1t} - b_{12}x_{2t}) + (\beta_2 + \beta_1 b_{12})x_{2t} + \epsilon_t \tag{66}
\]

\[
= \delta x_{2t} + u_t \tag{67}
\]

43
where as before, \( b_{12} \) is the parameter from the completing model, \( X_{1t} = b_{12}x_{2t} + u_t \), and the DGP \( M_0 \) implies,

\[
\delta = \beta_2 + \beta_1 b_{12}
\]  

(68)

The encompassing question is whether \( M_1 \) can explain the parameter \( \delta \) of model \( M_2 \). We proceed as if \( M_1 \) is a congruent model, so it is a reduction of the DGP where \( \beta_2 = 0 \). In other words, a sufficient condition for \( M_1 \) to explain the results of \( M_2 \) is for \( \beta_2 = 0 \), because this implies that \( \alpha = \beta_1 \) (from equation 64), and therefore \( M_1 \) coincides with \( M_0 \). Now, because \( M_0 \) can explain \( \delta \), so can \( M_1 \). Both models predict

\[
\delta_\alpha = \beta_2 + \beta_1 b_{12}
\]

\[
= 0 + \beta_1 b_{12}
\]

\[
= \alpha b_{12}
\]  

(69)

If instead, \( \beta_2 \neq 0 \), then \( M_1 \) fails to explain \( M_2 \). \( M_1 \)'s prediction of \( \delta \) is

\[
\delta_\alpha = \alpha b_{12}
\]  

(70)

which using \( \alpha \) from \( M_0 \) is

\[
\delta_\alpha = (\beta_1 + \beta_2 b_{21})b_{12}
\]  

(71)

while \( \delta \) from \( M_0 \) is

\[
\delta = \beta_2 + \beta_1 b_{12}
\]  

(72)
Thus, $\delta \neq \delta_\alpha$ unless $b_{12}b_{21} = 1$, which requires perfect correlation between $x_{2t}$ and $x_{1t}$, so that $M_1$ is isomorphic to $M_2$.

Formalizing Encompassing

We now have three different representations, $M_0$, $M_1$, and $M_2$. The theory of reduction tells us that $M_1$ and $M_2$ are nested in the DGP, so their parameters can be written as functions of those of the DGP. If we denote the parameters $(\beta_1, \beta_2, \sigma^2)$ of the DGP as $\theta$, then we can write $\alpha = \alpha[\theta]$ and $\delta = \delta[\theta]$. For example, equation (68) above. When $M_1$ is assumed to be the DGP, then $M_2$ is nested within $M_1$ despite its appearance, and $\delta$ must be a function of $\alpha$. Write this as

$$\delta_\alpha = \phi_{21}[\alpha]$$

(73)

where $\phi_{21}$ denotes a mapping from the parameter space of $M_1$ to that of $M_2$. So $\delta_\alpha$ is what $M_1$ predicts $M_2$ should find as the value of $\delta$.

Tests of encompassing are based on the difference between what $M_2$ finds, and what $M_1$ predicts $M_2$ should find:

$$\psi_\delta = \delta - \phi_{21}[\alpha]$$

(74)

- Model 1 encompasses Model 2, denoted $M_1 \epsilon M_2$ with respect to $\delta$ if $\psi_\delta = 0$.

- If $M_1 \epsilon M_2$, then either $M_2 \epsilon^c M_1$, or the two models are equivalent.
Consider the two models

\begin{align*}
M_1: \quad y &= X\beta + u \quad u \sim N[0, \sigma^2 I] \tag{75} \\
M_2: \quad y &= Z\gamma + v \quad u \sim N[0, \tau^2 I] \tag{76}
\end{align*}

where \( \alpha = (\beta, \sigma^2) \) is a \((k + 1) \times 1\) parameter vector.

\begin{align*}
M_2^*: \quad y &= Z\gamma + v \quad u \sim N[0, \tau^2 I] \tag{77}
\end{align*}

where \( \delta = (\gamma, \tau^2) \) is a \((n + 1) \times 1\) parameter vector.

Assume that \( Z \) and \( X \) contain distinct variables, so that neither is a linear transformation of a subset of the other.

Now consider the relationship between \( X \) and \( Z \) in terms of the completing model \( M_c \) which introduces the auxiliary equation:

\begin{align*}
M_c: \quad X &= Z\pi + w, \quad w \sim N[0, \Omega]. \tag{78}
\end{align*}

Now substitute (78) into (75). A researcher who believes that \( M_1 \) and \( M_c \) are both correct, would derive the predicted form of \( M_2^* \) as:

\begin{align*}
M_2^*: \quad y &= Z\Pi\beta + u + w\beta \\
y &= Z\gamma_\alpha + v, \quad \gamma_\alpha = \Pi\beta \tag{79}
\end{align*}
From $M_1$ and $M_c$, $\gamma_\alpha = \phi_{21}[\alpha^*]$, where $\alpha^* = (\alpha, \Pi, \Omega)$. $M_2^*$ is a linear regression equation relating $Y$ to $Z$ which is anticipated under $M_1$. $M_1$ also predicts

$$v = u + w\beta \quad \Rightarrow \quad \tau^2 = \sigma^2 + \beta'\Omega\beta \quad (80)$$

Both $\gamma$ and $\tau^2$ may be estimated either directly or indirectly. For instance, $\gamma$ can be estimated by regressing $Y$ on $Z$, or by regressing $Y$ on $X$ and $X$ on $Z$ to produce $\gamma_\alpha$.

Under $H_0 : M_1 \varepsilon M_2$, then $\gamma_\alpha = \Pi\beta$, and $\tau^2 = \sigma^2 + \beta'\Omega\beta$. So we test encompassing by testing

$$H_0 : \psi_\alpha = 0 \quad \text{where} \quad \psi_\alpha = \gamma - \gamma_\alpha \quad \quad \quad (81)$$

$$\quad \quad = \alpha - \Pi\beta$$

or

$$H_0 : \psi_{\tau^2} = 0 \quad \text{where} \quad \psi_{\tau^2} = \tau^2 - \sigma^2 - \beta'\Omega\beta \quad (82)$$

PcGive conducts a variety of encompassing tests. The test for variance encompassing (82) is referred to as the Cox test and is due to Cox (1961). The Joint Model F-test in PcGive is similar to (82) but it actually tests what is called parsimonious encompassing. It forms the minimal nesting model, $M_m = M_1U M_2^\perp$, or in our case $M_m : y = X\beta + Z\gamma + \epsilon_i$ and then tests the hypothesis that delivers the model hypothesized to nest the rival model. To conduct encompassing tests in PcGive, first estimate the two competing models. Then from the Model Menu select Encompassing. The null hypothesis is that one model encompasses the other, so we reject the null if the reported marginal significance level is less than or equal
to our chosen significance level, say ten percent. (See Doornik and Hendry 1996, p. 150, and p. 243.)

**Forecast Encompassing**


Let \( \hat{y} \) be the 1-step-ahead forecast of \( y \) from model \( M_1 \), and \( \tilde{y} \) be the 1-step-ahead forecast of \( y \) from model \( M_2 \). The pooled forecast is given by:

\[
\bar{y} = (1 - \omega)\hat{y} + \omega\tilde{y} \tag{83}
\]

If the minimum mean squared forecast error (MMSFE) value of \( \omega \) is neither zero or unity, then both forecasts are useful in predicting \( y \), but neither is sufficient. In other words, if we pick \( \omega \) to minimize the squared forecast error of \( \bar{y} \), and \( \omega \) is neither zero or one, then combining the two forecasts improves the predictive power of \( \bar{y} \). Unfortunately, combining forecasts is not a very useful strategy for improving forecasts. Two failures are not likely to systematically combine to form a success.

The test for forecast encompassing is the same as the test for conditional efficiency first proposed in Nelson (1972), and Granger and Newbold (1973). A forecast is conditionally efficient if the variance of the combination forecast error, \( \bar{y} \), is not significantly less than that
of the original forecast error. Let \( y - \hat{y} = \hat{u} \) be the forecast error from \( M_1 \). From equation (83), we can write the forecast error of the combination forecast as,

\[
y - \bar{y} = y - \hat{y} + \omega(\hat{y} - \bar{y})
\]

or

\[
\bar{u} = \hat{u} - \omega(\hat{u} - \bar{u}) \tag{84}
\]

so \( V[\bar{u}] < V[\hat{u}] \) only if \( \omega \neq 0 \). If \( \omega \neq 0 \), then the difference between the forecast errors of the two models helps explain the forecast errors of \( M_1 \), so that \( M_1 \) could be improved by incorporating some of the features of \( M_2 \). The forecast encompassing test is conducted by testing \( H_0 : \alpha = 0 \) in,

\[
y - \hat{y} = \alpha(\hat{y} - \bar{y}) + \epsilon_t \tag{85}
\]

4.4 Congruent Models

A model which satisfies all six of these criteria is described as being congruent with the evidence.