1.3 Geometric (or Koyck) Distributed Lag Models

The geometric distributed lag model restricts the coefficients to decay at a geometric rate

$$\beta_i = \beta(1-\lambda)\lambda^i \tag{4}$$

for $i = 0, 1, 2, ...$ Substitution of (4) into (1) gives

$$y_t = \alpha + \beta(1-\lambda)x_t + \beta(1-\lambda)\lambda x_{t-1} + \beta(1-\lambda)\lambda^2 x_{t-2} + \cdots + \epsilon_t. \tag{5}$$

Lag equation (5) once and multiply by λ

$$\lambda y_{t-1} = \lambda\alpha + \beta(1-\lambda)\lambda x_{t-1} + \beta(1-\lambda)\lambda^2 x_{t-2} + \beta(1-\lambda)\lambda^3 x_{t-3} + \cdots + \lambda\epsilon_{t-1}. \tag{6}$$

Subtract (6) from (5)

$$\begin{aligned} y_t &= \alpha(1-\lambda) + \lambda y_{t-1} + \beta(1-\lambda)x_t + (\epsilon_t - \lambda\epsilon_{t-1}) \\ &= \alpha^* + \lambda y_{t-1} + \beta^* x_t + \mu_t. \end{aligned}$$

Advantages.

1. Can greatly reduce the number of parameters to estimate.
2. Reduces multicollinearity.

Disadvantages.

1. May be imposing incorrect (geometric) functional form on the coefficients.
2. μ_t is going to be autocorrelated (unless ϵ_t happened to follow an AR(1) process with coefficient λ).
3. The $corr(y_{t-1}, \mu_t) \neq 0$, so instrumental variable estimation will be necessary.

2 Univariate Stationary Processes

Univariate time series models can be used for generating forecasts. In fact, simple time series models often dominate large-scale theoretic macroeconomic models in terms of forecasting the future.

Recall, a stochastic time series process $\{y_t\}_{t=1}^{\infty}$ is weakly covariance stationary (WCS) if

- $\mu = E(y_t)$ is a finite constant $\forall t$

- $\lambda_0 = var(y_t)$ is a finite constant $\forall t$
- $\lambda_k = cov(y_t, y_{t-k})$ is finite and depends on k not t.

2.1 Autoregressive-Moving Average Processes

Consider the following ARMA(p,q) process:

$$y_t = \mu + \underbrace{\phi_1 y_{t-1} + \cdots + \phi_p y_{t-p}}_{\text{AR part}} + \epsilon_t + \underbrace{\theta_1 \epsilon_{t-1} + \cdots + \theta q \epsilon_{t-q}}_{\text{MA part}}. \tag{7}$$

Now consider some commonly used examples of (7).

2.1.1 AR(1)

A first-order autoregressive, AR(1), process takes the form

$$y_t = \phi_1 y_{t-1} + \epsilon_t \tag{8}$$

where $|\phi| < 1$ and ϵ_t is mean-zero white noise. Repeated substitutions produces the MA(∞) process

$$y_t = \epsilon_t + \phi_1 \epsilon_{t-1} + \phi_1^2 \epsilon_{t-2} + \phi_1^3 \epsilon_{t-3} + \cdots$$

The first and second moments of y_t are

$$\begin{aligned}
\mu &= E(y_t) = 0 \\
\lambda_0 &= var(y_t) = \sigma^2(1 + \phi_1^2 + \phi_1^4 + \cdots) = \frac{\sigma^2}{(1 - \phi_1^2)} \\
\lambda_k &= cov(y_t, y_{t-k}) = E\left[(\epsilon_t + \phi_1 \epsilon_{t-1} + \phi_1^2 \epsilon_{t-2} + \cdots)(\epsilon_{t-k} + \phi_1 \epsilon_{t-k-1} + \phi_1^2 \epsilon_{t-k-2} + \cdots)\right] \\
&= \sigma^2(\phi_1^k + \phi_1^{k+2} + \phi_1^{k+4} + \cdots) \\
&= \frac{\sigma^2 \phi_1^k}{(1 - \phi_1^2)} = \phi_1^k \lambda_0.
\end{aligned}$$

Therefore, the complete autocovariance function is

$$\lambda_k = \phi_1^k \lambda_0$$

for $k = 0, 1, 2, ...$ The autocorrelation function (ACF) is

$$\rho_k = \frac{\lambda_k}{\lambda_0} = \phi_1^k$$

for $k = 0, 1, 2, ...$ The partial autocorrelation function (PACF) is given by the last coefficients in a linear projection of y_t on its past values. In general, the k^{th} population autocorrelation coefficient (ρ_k^*) is taken from

$$y_t = \phi_1^{(k)} y_{t-1} + \cdots + \phi_{k-1}^{(k)} y_{t-(k-1)} + \rho_k^* y_{t-k} + \epsilon_t.$$

For the AR(1) process, the PACF is

$$\rho_k^* = \{ \begin{array}{l} \phi_1^k, \ k = 0, 1 \\ 0, \ k > 1. \end{array}$$

2.1.2 MA(1)

A first-order moving average, MA(1), process takes the form

$$y_t = \epsilon_t + \theta_1 \epsilon_{t-1}$$

where ϵ_t is mean-zero white noise. The first and second moments are

$$\begin{aligned} \mu &= E(y_t) = 0 \\ \lambda_0 &= var(y_t) = \sigma^2(1 + \theta_1^2) \\ \lambda_1 &= cov(y_t, y_{t-1}) = E[(\epsilon_t + \theta_1 \epsilon_{t-1})(\epsilon_{t-1} + \theta_1 \epsilon_{t-2})] = \theta_1 \sigma^2. \end{aligned}$$

The ACF is

$$\rho_k = \{ \begin{array}{l} 1, \ k = 0 \\ \theta_1/(1 + \theta_1^2), \ k = 1 \\ 0, \ k > 1. \end{array}$$

The PACF is found by inverting the MA process in the same fashion we inverted the AR(1) process to obtain an MA(∞) process:

$$y_t = \theta(L)\epsilon_t$$

where $\theta(L) = 1 + \theta_1 L$. This implies that the AR representation of the process is

$$\theta^{-1}(L) y_t = \epsilon_t$$

where $\theta^{-1}(L) = 1 - \theta_1 L + \theta_1^2 L - \theta_1^3 L + \cdots$ will exist if the eigenvalues of $\theta(z)$ lie outside the unit circle. Therefore,

$$y_t = \theta_1 y_{t-1} - \theta_1^2 y_{t-2} + \theta_1^3 y_{t-3} - \cdots + \epsilon_t. \tag{9}$$

Using (9), we can see that the PACF is therefore

$$\rho_k^* = (-1)^{k+1}\theta_1^k$$

for $k = 0, 1, 2, ...$

2.1.3 ARMA(1,1)

An ARMA(1,1) process takes the form

$$y_t = \phi_1 y_{t-1} + \epsilon_t + \theta_1 \epsilon_{t-1} \tag{10}$$

where the conditions for stationarity hold (i.e., the roots of $\phi(z) = 1 - \phi_1 z$ lie outside the unit circle). Multiply equation (10) by y_t, y_{t-1} and y_{t-2}:

$$\begin{aligned} y_t^2 &= \phi_1 y_t y_{t-1} + y_t \epsilon_t + \theta_1 y_t \epsilon_{t-1} \\ y_{t-1} y_t &= \phi_1 y_{t-1}^2 + y_{t-1} \epsilon_t + \theta_1 y_{t-1} \epsilon_{t-1} \\ y_{t-2} y_t &= \phi_1 y_{t-2} y_{t-1} + y_{t-2} \epsilon_t + \theta_1 y_{t-2} \epsilon_{t-1} \end{aligned}$$

and take expectations:

$$\begin{aligned} \lambda_0 &= \phi_1 \lambda_1 + \sigma^2 + \theta_1(\phi_1 \sigma^2 + \theta_1 \sigma^2) \\ &= \phi_1 \lambda_1 + \sigma^2(1 + \theta_1(\phi_1 + \theta_1)) \\ \lambda_1 &= \phi_1 \lambda_0 + \theta_1 \sigma^2 \\ \lambda_k &= \phi_1 \lambda_{k-1} \end{aligned}$$

for $k = 2, 3, ...$ These equations are called Yule-Walker equations and can be used to solve for the ACF (as well as the parameters in terms of the autocorrelations).

2.2 Box-Jenkins Approach to Modeling

Box and Jenkins' methodology is supported by Wold's Theorem, which states that any stationary stochastic process can be decomposed into a linearly deterministic part, $E(y_t|y_{t-1}, ..., y_{t-p})$, and a linearly indeterministic part, $\sum_{i=0}^{\infty} \pi_i \epsilon_{t-i}$. Choosing the correct ARMA process using the Box-Jenkins approach is as much art as science. Here are the general steps:

1. Check for stationarity, if not, induce.

2. Identify the AR and MA orders from the ACF and PACF.

3. Estimate the parameters and perform diagnostic testing.

4. Forecast.

The best way to understand how the Box-Jenkins methodology works is to see it in action. Consider housing starts in the United States. The final ARMA model is

$$(1-\phi_1 L)(1-\phi_1^s L^{12})y_t = (1-\theta_1 L)(1-\theta_1^s L^{12})\epsilon_t.$$

3 Vector Stationary Processes

Consider the vector autoregressive (VAR) process with p lags

$$y_t = \Phi_1 y_{t-1} + \Phi_2 y_{t-2} + \cdots + \Phi_p y_{t-p} + \epsilon_t$$

where $y_t = (y_{1t}, y_{2t}, ..., y_{mt})'$ and $\epsilon_t \sim iid(0, \Omega)$. Rewrite as a VAR(1) process:

$$\xi_t = F\xi_{t-1} + v_t \tag{11}$$

where $v_t \sim iid(0, Q)$ and

$$\xi_t = \begin{bmatrix} y_t \\ y_{t-1} \\ \vdots \\ y_{t-p+1} \end{bmatrix}_{mp\times 1} \quad F = \begin{bmatrix} \Phi_1 & \Phi_2 & \cdots & \Phi_{p-1} & \Phi_p \\ I_m & 0 & & 0 & 0 \\ 0 & I_m & & 0 & 0 \\ \vdots & & \ddots & & \vdots \\ 0 & 0 & & I_m & 0 \end{bmatrix}_{mp\times mp} \quad v_t = \begin{bmatrix} \epsilon_t \\ 0 \\ \vdots \\ 0 \end{bmatrix}_{mp\times 1}.$$

The vector y_t is stationary if the eigenvalues of F lie inside the unit circle. Assuming y_t is stationary, the autocovariances can be found by (post) multiplying (11) by ξ_t' and taking expected values:

$$\begin{aligned} E(\xi_t \xi_t') &= E[(F\xi_{t-1} + v_t)(F\xi_{t-1} + v_t)'] \\ &= FE[\xi_{t-1}\xi_{t-1}']F' + Q. \end{aligned}$$

This implies

$$\Sigma_0 = F\Sigma_0 F' + Q.$$

A closed-form solution can be found by using the vec operator

$$\begin{aligned} vec(\Sigma_0) &= (F \otimes F)vec(\Sigma_0) + vec(Q) \\ &\Rightarrow \quad vec(\Sigma_0) = [I - (F \otimes F)]^{-1} vec(Q). \end{aligned}$$

Higher autocovariances are given by

$$\Sigma_j = F\Sigma_{j-1} = F^j \Sigma_0$$

for $j = 1, 2, ...$ Next, we look at some common uses of VARs.

3.0.1 Granger Causality

Variable y_2 is said to Granger-cause y_1 if lagged values of y_2 explain the variation in y_1, above and beyond the lagged values of y_1. Put differently, if equation-by-equation OLS estimation ($p = 1$) of

$$\begin{bmatrix} y_{1t} \\ y_{2t} \end{bmatrix} = \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix} + \begin{bmatrix} \phi_{11} & \phi_{12} \\ \phi_{21} & \phi_{22} \end{bmatrix} \begin{bmatrix} y_{1,t-1} \\ y_{2,t-1} \end{bmatrix} + \begin{bmatrix} \epsilon_{1t} \\ \epsilon_{2t} \end{bmatrix}$$

shows that $\phi_{12} = 0$ (e.g., using a standard F test), then y_2 does not Granger-cause y_1. Equation-by-equation OLS is efficient because this is essentially a SUR system with identical regressors. It is important to recognize that an absence of Granger causality is not the same thing as exogeneity, as Granger causality only deals with lagged values. It is indeed possible that y_{1t} and y_{2t} are contemporaneously correlated even after accounting for lagged values of y_{1t}.

3.0.2 Impulse Response Functions

For example, if we were to perturb ϵ_{1t} and trace out the effects on y_{1t} over time, this would be called an impulse response function (IRF). A common problem is that $corr(\epsilon_{1t}, \epsilon_{2t}) \neq 0$. The standard solution is to orthogonalize the innovations. Let $u_t = P\epsilon_t$ where $P\Omega P' = I$ and P is a lower triangular matrix. This is referred to as the Choleski decomposition and produces

$$Py_t = P\mu + P\Phi_1 y_{t-1} + u_t.$$

One disadvantage of the Choleski decomposition is that the ordering of the variables in the vector y_t matters.

3.0.3 Forecasting and Variance Decomposition (VDC)

The optimal forecast of ξ_{t+s} is

$$\hat{\xi}_{t+s} = E(\xi_{t+s}|\xi_t, ..., \xi_1) = F^s \xi_t$$

with forecast error

$$\begin{aligned} e_s &= \xi_{t+s} - \hat{\xi}_{t+s} = F^s\xi_t + v_{t+s} + Fv_{t+s-1} + \cdots + F^{s-1}v_{t+1} - F^s\xi_t \\ &= v_{t+s} + Fv_{t+s-1} + \cdots + F^{s-1}v_{t+1} \end{aligned}$$

and variance

$$\Sigma(s) = Q + FQF' + \cdots + F^{s-1}Q(F^{s-1})'.$$

Substitute in $Q = P_*^{-1}(P_*^{-1})'$, the redimensioned version of the Choleski decomposition,

$$\begin{aligned} \Sigma(s) &= P_*^{-1}(P_*^{-1})' + FP_*^{-1}(P_*^{-1})'F' + \cdots + F^{s-1}P_*^{-1}(P_*^{-1})'(F^{s-1})' \\ &= P_*^{-1}(P_*^{-1})' + (FP_*^{-1})(FP_*^{-1})' + \cdots + (F^{s-1}P_*^{-1})(F^{s-1}P_*^{-1})'. \end{aligned} \tag{12}$$

Equation (12) can be used to decompose the forecast error s periods ahead into the parts due to each orthogonal impulse.

3.0.4 VAR Application

Consider the following structural bivariate system

$$X_t = aY_t + bY_{t-1} + cX_{t-1} + e_t \tag{13}$$

$$Y_t = fX_t + gX_{t-1} + hY_{t-1} + v_t \tag{14}$$

where $var(\epsilon_t) = var[(e_t, v_t)'] = I$. The standard identification conditions apply in order to identify the structural parameters. For example, $c = h = 0$ would be sufficient to identify the remaining parameters from the reduced form. The reduced-form equations are

$$\begin{aligned} X_t &= (1-af)^{-1}\left[(ag+c)X_{t-1} + (ah+b)Y_{t-1} + (av_t + e_t)\right] \\ Y_t &= (1-af)^{-1}\left[(fc+g)X_{t-1} + (fb+h)Y_{t-1} + (fe_t + v_t)\right] \end{aligned}$$

which can be simplified to

$$\begin{aligned} \begin{bmatrix} X_t \\ Y_t \end{bmatrix} &= \begin{bmatrix} \pi_{11} & \pi_{12} \\ \pi_{21} & \pi_{22} \end{bmatrix} \begin{bmatrix} X_{t-1} \\ Y_{t-1} \end{bmatrix} + (1-af)^{-1} \begin{bmatrix} 1 & a \\ f & 1 \end{bmatrix} \begin{bmatrix} e_t \\ v_t \end{bmatrix} \\ &= \Pi \begin{bmatrix} X_{t-1} \\ Y_{t-1} \end{bmatrix} + A_0\epsilon_t = \Pi \begin{bmatrix} X_{t-1} \\ Y_{t-1} \end{bmatrix} + \begin{bmatrix} w_t^x \\ w_t^y \end{bmatrix} \end{aligned}$$

where $var(A_0\epsilon_t) = \Omega$. Clearly the $corr(w_t^x, w_t^y) \neq 0$. One method to identify the structural shocks (e_t and v_t) is through the Choleski decomposition, which given the lower triangular nature of P, imposes that $a = 0$. The Choleski decomposition, switching the order of X and Y, imposes $f = 0$.

Repeated substitutions of the VAR(1) process above will produce

$$\begin{bmatrix} X_t \\ Y_t \end{bmatrix} = \Pi^s \begin{bmatrix} X_{t-s} \\ Y_{t-s} \end{bmatrix} + A_0 \textstyle\sum_{i=0}^{s-1} \Pi^i \epsilon_{t-i}.$$

Letting $s \to \infty$, and assuming that $(X_t, Y_t)'$ is a stationary vector process, produces the MA(∞) process

$$\begin{bmatrix} X_t \\ Y_t \end{bmatrix} = A_0(\epsilon_t + \Pi\epsilon_{t-1} + \Pi^2\epsilon_{t-2} + \cdots).$$

Imposing the Choleski decomposition $P\Omega P' = I$ (implying that $PA_0\epsilon_t = u_t$) gives

$$\begin{bmatrix} X_t \\ Y_t \end{bmatrix} = P^{-1}u_t + \Pi P^{-1}u_{t-1} + \Pi^2 P^{-1}u_{t-2} + \cdots$$

We can estimate $\Omega = A_0A_0' = P^{-1}(P^{-1})'$, which provides three pieces of information. And although there are four unknown elements in A_0, the Choleski decomposition ($A_0 = P^{-1}$) imposes the single restriction $a = 0$, so as to exactly identify the structural shocks.

Numerical Example (Granger causality, Choleski decomposition, IRF and VDC). Consider the VAR

$$\begin{bmatrix} X_{t+1} \\ Y_{t+1} \end{bmatrix} = \begin{bmatrix} 1/2 & 0 \\ 3/4 & 1/2 \end{bmatrix} \begin{bmatrix} X_t \\ Y_t \end{bmatrix} + \begin{bmatrix} w_{t+1}^x \\ w_{t+1}^y \end{bmatrix} \tag{15}$$

with

$$var[(w_{t+1}^x, w_{t+1}^y)'] = \begin{bmatrix} 2 & 3/2 \\ 3/2 & 6 \end{bmatrix} = \Omega.$$

Now let's examine the implications for Granger causality, the Choleski decomposition, IRFs and VDCs.

1. Granger causality. From the structure of (15), we know that $X \xrightarrow{GC} Y$ but $Y \overset{GC}{\nrightarrow} X$.

2. Choleski decomposition.

$$\begin{bmatrix} 2 & 3/2 \\ 3/2 & 6 \end{bmatrix} = \begin{bmatrix} P_{11} & 0 \\ P_{12} & P_{22} \end{bmatrix} \begin{bmatrix} P_{11} & P_{12} \\ 0 & P_{22} \end{bmatrix} = \begin{bmatrix} P_{11}^2 & P_{11}P_{12} \\ P_{11}P_{12} & P_{12}^2 + P_{22}^2 \end{bmatrix}$$

This produces three equations in three unknowns

$$\begin{aligned} P_{11}^2 &= 2 \\ P_{11}P_{12} &= 3/2 \\ P_{12}^2 + P_{22}^2 &= 6. \end{aligned}$$

Solving this system gives $P_{11} = \sqrt{2}$, $P_{12} = 3\sqrt{2}/4$ and $P_{22} = \sqrt{39}/(2\sqrt{2})$. Therefore,

$$P^{-1} = \begin{bmatrix} \sqrt{2} & 0 \\ 3\sqrt{2}/4 & \sqrt{39}/(2\sqrt{2}) \end{bmatrix}.$$

3. Impulse response function. Start at time t and set $X_t = Y_t = 0$. The impulse response functions for X and Y are

$$\begin{aligned} \begin{bmatrix} X_{t+1} \\ Y_{t+1} \end{bmatrix} &= \Pi \begin{bmatrix} 0 \\ 0 \end{bmatrix} + \begin{bmatrix} \sqrt{2} & 0 \\ 3\sqrt{2}/4 & \sqrt{39}/(2\sqrt{2}) \end{bmatrix} \begin{bmatrix} 1 \\ 0 \end{bmatrix} = \begin{bmatrix} \sqrt{2} \\ 3\sqrt{2}/4 \end{bmatrix} = \begin{bmatrix} 1.414 \\ 1.061 \end{bmatrix}. \\ \begin{bmatrix} X_{t+2} \\ Y_{t+2} \end{bmatrix} &= \begin{bmatrix} 1/2 & 0 \\ 3/4 & 1/2 \end{bmatrix} \begin{bmatrix} \sqrt{2} \\ 3\sqrt{2}/4 \end{bmatrix} = \begin{bmatrix} \sqrt{2}/2 \\ 9\sqrt{2}/8 \end{bmatrix} = \begin{bmatrix} 0.707 \\ 1.591 \end{bmatrix}. \\ \begin{bmatrix} X_{t+3} \\ Y_{t+3} \end{bmatrix} &= \begin{bmatrix} 1/2 & 0 \\ 3/4 & 1/2 \end{bmatrix} \begin{bmatrix} \sqrt{2}/2 \\ 9\sqrt{2}/8 \end{bmatrix} = \begin{bmatrix} \sqrt{2}/4 \\ 15\sqrt{2}/16 \end{bmatrix} = \begin{bmatrix} 0.354 \\ 1.326 \end{bmatrix}. \end{aligned}$$

4. Variance decomposition. The s-step ahead forecast error is

$$e_s = v_{t+s} + \Pi v_{t+s-1} + \cdots + \Pi^{s-1} v_{t+1}$$

with variance

$$\Sigma(s) = P^{-1}(P^{-1})' + \Pi[P^{-1}(P^{-1})']\Pi' + \cdots + \Pi^{s-1}[P^{-1}(P^{-1})'](\Pi^{s-1})'.$$

Let $s = 1$.

$$\begin{aligned}
\Sigma(1) &= P^{-1}(P^{-1})' = \begin{bmatrix} 2 & 3/2 \\ 3/2 & 6 \end{bmatrix} \\
\Sigma^1(1) &= P^{-1}\begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}(P^{-1})' = \begin{bmatrix} \sqrt{2} & 0 \\ 3\sqrt{2}/4 & \sqrt{39}/(2\sqrt{2}) \end{bmatrix}\begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}\begin{bmatrix} \sqrt{2} & 3\sqrt{2}/4 \\ 0 & \sqrt{39}/(2\sqrt{2}) \end{bmatrix} = \begin{bmatrix} 2 & 3/2 \\ 3/2 & 9/8 \end{bmatrix} \\
\Sigma^2(1) &= P^{-1}\begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}(P^{-1})' = \begin{bmatrix} \sqrt{2} & 0 \\ 3\sqrt{2}/4 & \sqrt{39}/(2\sqrt{2}) \end{bmatrix}\begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}\begin{bmatrix} \sqrt{2} & 3\sqrt{2}/4 \\ 0 & \sqrt{39}/(2\sqrt{2}) \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ 0 & 39/8 \end{bmatrix}.
\end{aligned}$$

These numbers can be interpreted as follows:

- $(\Sigma^1(1))_{11}/(\Sigma(1))_{11} = 2/2 = 100\%$ of one-step ahead variance in X is explained by X orthogonal shock.
- $(\Sigma^2(1))_{11}/(\Sigma(1))_{11} = 0/2 = 0\%$ of one-step ahead variance in X is explained by Y orthogonal shock.
- $(\Sigma^1(1))_{22}/(\Sigma(1))_{22} = (9/8)/6 \simeq 19\%$ of one-step ahead variance in Y is explained by X orthogonal shock.
- $(\Sigma^2(1))_{22}/(\Sigma(1))_{22} = (39/8)/6 \simeq 81\%$ of one-step ahead variance in Y is explained by Y orthogonal shock.

4 Nonstationary Processes

Nonstationarity can come in various forms, however, the two most common involve linear and stochastic trends. Nonstationary series that can be made stationary through differencing are often said to be integrated.

4.0.5 Trend Stationary (TS) Processes

One example of a TS process is

$$y_t = \mu + \beta t + \epsilon_t, \tag{16}$$

which is said to have a unit root because it can be rewritten as

$$(1 - L)y_t = \beta + (\epsilon_t - \epsilon_{t-1})$$

where the characteristic root of the lag term $(1-z)$ is one. First differencing the data will induce stationarity, but regressing y_t on a time trend is preferable.

4.0.6 Difference Stationary (DS) Processes

One example of a DS process is

$$y_t = \mu + y_{t-1} + \epsilon_t, \tag{17}$$

which also has a unit root and can be rewritten as

$$(1 - L)y_t = \mu + \epsilon_t.$$

The process for y_t is referred to as a random walk with drift μ. The term random walk refers to the fact that y_t does not revert to a long-run mean but is free to roam about with gradual drift μ. The big difference (so to speak) between the TS and DS formulations is the error term. In the TS specification, ϵ_t has a temporary effect; while in the DS specification, ϵ_t has a permanent effect.

4.0.7 Unit Root Tests

To test for a unit root, the naive way to proceed is to run the regression

$$y_t = \mu + \alpha y_{t-1} + \epsilon_t$$

and test to see if $\alpha = 1$. Unfortunately, however, $\hat{\alpha}_{OLS}$ is known to be downwardly biased because under the null (H_0: $\alpha = 1$), the series is not stationary.

An alternative approach, advocated by Dickey and Fuller, is to nest equations (16) and (17) according to

$$y_t = \mu + \beta(1 - \alpha)t + \alpha y_{t-1} + u_t$$

and then subtract y_{t-1}:

$$\begin{aligned} (1 - L)y_t &= \mu + \beta^* t + (\alpha - 1)y_{t-1} + u_t \\ &= \mu + \beta^* t + \gamma y_{t-1} + u_t. \end{aligned} \tag{18}$$

The Dickey-Fuller test then estimates (18) by OLS and tests the hypothesis

$$\begin{aligned} H_0 &: \gamma = \beta^* = 0 \text{ (i.e., } \alpha = 1) \\ H_A &: H_0 \text{ false} \end{aligned}$$

by forming the statistic

$$DF = \hat{\gamma}/se(\hat{\gamma})$$

and referring to the critical values of the limiting distribution (see Greene, Table 20.4). The DF statistic does not have a t distribution because under the null hypothesis, the process is not stationary. Augmented DF tests are available which add first-differenced lags to equation (18) so as to account for serial correlation. Other tests (e.g., Phillips and Perron test) have been suggested to improve on the small-sample performance of the DF test.

5 Cointegration

Let x_t and y_t be two first-order integrated, $I(1)$, variables. In other words, each variables has a unit root. Typically, linear combinations of x_t and y_t will also be $I(1)$. If, however, $x_t + \lambda\ y_t$ is $I(0)$, then x_t and y_t are said to be cointegrated or share a common trend.

5.0.8 An Example.

Consider the example of purchasing power parity (PPP)

$$\begin{aligned} \ln(\epsilon_t^*) &= \alpha_0 + \alpha_1 t + \nu_t^* \\ \ln(\epsilon_t) &= \beta_0 + \beta_1 t + \nu_t \end{aligned}$$

where $\epsilon_t^* \equiv$ Japan/US real exchange rate and $\epsilon_t \equiv$ UK/US real exchange rate. Combining the two gives

$$\ln(\epsilon_t^*) - \ln(\epsilon_t) = (\alpha_0 - \beta_0) + (\alpha_1 - \beta_1)t + (\nu_t^* - \nu_t).$$

If PPP holds, $\alpha_0 = \beta_0$ and $\alpha_1 = \beta_1$ in the long run; ϵ_t^* and ϵ_t are cointegrated.

5.0.9 Testing for Cointegration

If $\epsilon_t = x_t + \lambda\ y_t$ is $I(0)$, then x_t and y_t are cointegrated. Regress y_t on x_t using OLS, form the residuals $\hat{\epsilon}_t$ and test for a unit root (because we are using $\hat{\epsilon}_t$ rather than ϵ_t, we use MacKinnon's critical values, not the DF critical values). The estimates of λ will be superconsistent, meaning they will converge in probability to λ at a faster rate than standard estimates.

5.0.10 Error-Correction Models

When modeling two variables $z_t = (x_t, y_t)'$ that are cointegrated, the appropriate framework is the error-correction model

$$\Delta z_t = \Phi \Delta z_{t-1} + \Pi z_{t-1} + \epsilon_t$$

where the terms in Πz_{t-1} are the error-correction terms. This can be written more intuitively as

$$\begin{aligned}\begin{bmatrix}\Delta x_t \\ \Delta y_t\end{bmatrix} &= \Phi \begin{bmatrix}\Delta x_{t-1} \\ \Delta y_{t-1}\end{bmatrix} + \begin{bmatrix}\rho_x(x_{t-1} + \lambda y_{t-1}) \\ \rho_y(x_{t-1} + \lambda y_{t-1})\end{bmatrix} + \begin{bmatrix}\epsilon_t^x \\ \epsilon_t^y\end{bmatrix} \\ &= \Phi \begin{bmatrix}\Delta x_{t-1} \\ \Delta y_{t-1}\end{bmatrix} + \begin{bmatrix}\pi_{11}x_{t-1} + \pi_{12}y_{t-1} \\ \pi_{21}x_{t-1} + \pi_{22}y_{t-1}\end{bmatrix} + \begin{bmatrix}\epsilon_t^x \\ \epsilon_t^y\end{bmatrix} \end{aligned} \tag{19}$$

where ρ_x and ρ_y (and hence the π's) are "speed of adjustment" parameters because they govern how fast x_t and y_t respond to deviations from their long-run equilibrium paths. For example, if $\lambda = -1$, then we would expect $\rho_x < 0$ and $\rho_y > 0$, with the magnitude of the coefficients governing the speed of the adjustment. Also, if x_t and y_t are cointegrated, we should expect the matrix Π to be of short rank, as can be seen from equation (19).

6 ARCH

Autoregressive conditional heteroscedasticity (ARCH) models are a reaction to "volatility clustering," the idea that in time-series settings, large and small errors often tend to be bunched in groups. The seminal paper in this area is Engle's (1982) article in *Econometrica*. An ARCH model takes the form

$$y_t = \beta_0 + \beta_1 x_t + \epsilon_t$$

where

$$E(\epsilon_t^2) = \sigma_t^2 = \alpha_0 + \alpha_1 \epsilon_{t-1}^2 + \cdots + \alpha_p \epsilon_{t-p}^2.$$

6.0.11 An Example

Consider an ARCH(1) model

$$\epsilon_t = \mu_t(\alpha_0 + \alpha_1 \epsilon_{t-1}^2)^{0.5}$$

where $\mu_t \sim iid(0, 1)$. The error term ϵ_t has the following first and second moments

$$\begin{aligned} E(\epsilon_t) &= 0 \\ var(\epsilon_t) &= \sigma^2 = \alpha_0/(1 - \alpha_1) \\ var(\epsilon_t|\epsilon_{t-1}) &= \sigma_t^2 = \alpha_0 + \alpha_1 \epsilon_{t-1}^2. \end{aligned}$$

6.0.12 Testing for ARCH

The standard procedure to test for ARCH is

Steps.

1. Regress y_t on x_t and save the residuals, e_t.
2. Run the OLS regression $e_t^2 = \alpha_0 + \alpha_1 e_{t-1}^2 + \cdots + \alpha_p e_{t-p}^2 + \nu_t$.
3. Test the hypothesis H_0: $\alpha_1 = \cdots = \alpha_p = 0$.
4. The test statistic is $LM = TR^2 \overset{asy}{\sim} \chi^2(p)$.

6.0.13 Estimation

Feasible GLS is possible using a two-step procedure (possibly iterated), but the most common approach is to assume normally distributed errors and use maximum likelihood, which is the (asymptotically) efficient estimator.

6.0.14 GARCH and ARCH-M

Several different variants of the original ARCH model have been suggested in the literature. Generalized ARCH (GARCH) takes the form

$$E(\epsilon_t^2) = \sigma_t^2 = \alpha_0 + \alpha_1 \epsilon_{t-1}^2 + \cdots + \alpha_p \epsilon_{t-p}^2 + \gamma_1 \sigma_{t-1}^2 + \cdots + \gamma_q \sigma_{t-q}^2.$$

The popular GARCH(1,1) model, in particular, is able to generate rich dynamics within a relatively parsimonious model. Another popular model is the ARCH-M model, which allows the conditional variance to affect the mean equation.

7 RATS Example

The time series methods examined above are applied to a vector system of U.S. money (M2) and income (real GDP). See RATS example 19-20.prg for details.