Methods of Economic Investigation I
Lent Term

Matthew Chesnes
The London School of Economics

March 22, 2002

1 Week 1: 14 Jan - 18 Jan

1.1 Time Series Models

- Observations will be correlated unlike cross-sectional data. Hence there are dependence issues that must be addressed.
- The observations are a sequence of random variables: a stochastic process.
- Example: $\epsilon_t \sim$ iid $(0, \sigma^2)$ with $t = 0, 1, 2, \ldots$ is a stochastic process.
- Other examples of stochastic processes:
  - White Noise: $y_t = \epsilon_t$.
  - Autoregression: (AR).
    $$AR(1) : y_t = \phi y_{t-1} + \epsilon_t.$$
    Clearly y_t is not iid. In general,
    $$AR(p) : y_t = \underbrace{\phi_1 y_{t-1} + \phi_2 y_{t-2} + \cdots + \phi_p y_{t-p}}_{p\ lags} + \epsilon_t.$$
    The AR process is a good model of economic shocks that occur in one period but continue to effect future periods with smaller influences (a multiplier effect).
  - Moving Average: (MA).
    $$MA(1) : y_t = \epsilon_t + \theta \epsilon_{t-1}.$$
    Here again, the observations are correlated across time. Note that shocks in period t affect y_t and y_{t+1} but NO further. This can also easily be extended to a $MA(p)$ process.
  - Autoregressive Moving Average Process: ARMA(p,q).
    $$y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + \cdots + \phi_p y_{t-p} + \epsilon_t + \theta_1 \epsilon_{t-1} + \cdots + \theta_q \epsilon_{t-q}.$$
    This is simply a difference equation with a stochastic error term.

1.2 Erdogicity

- Ergodicity is a property of stochastic processes. The process is called "Ergodic" if it satisfies the following two properties:
  - Stationarity: Moments are constant over time including cross moments such as $Var(y_t, y_{t-1})$. Thus means, variances, covariances, etc. are constant over time. In other words, $E[y_t]$, $Var(y_t)$, $Cov(y_t, y_{t-s})$ are all independent of time.
  - Limited Memory: As $s \to \infty$, $Cov(y_t, y_{t-s}) \to 0$. The correlation between two observations s periods apart goes to zero as s goes to infinity.

- Thus the two above properties imply ergodicity. Most series have limited memory, so ergodicity and stationarity can almost be used interchangably.
- However, most economic time series are non-stationary because they grow over time. Thus, by looking at deviations from a trend line might yield stationarity.
- The Ergodic Theorem: If we have y_t, $t = 1, 2, \ldots$ and y_t is ergodic with mean, μ, then:

$$plim \frac{1}{T} \sum_{t=1}^{T} y_t = \mu.$$

  Note that having y_t be *iid* also would give us this, but all we really need is ergodicity.

- If y_t and z_t are both ergodic, then:

$$ay_t + bz_t \text{ is ergodic.}$$

  Or more generally,

$$f(y_t, z_t) \text{ is ergodic if } f \text{ is continuous.}$$

1.3 The Moving Average Process

- Consider an $MA(1) : y_t = \epsilon_t + \theta\epsilon_{t-1}$. Thus,

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

- Since y_t has mean 0, then $Var(y_t) = E[y_t^2]$. Thus,

$$E[y_t^2] = E[(\epsilon_t + \theta\epsilon_{t-1})^2].$$

$$= E[\epsilon_t^2 + \theta^2\epsilon_{t-1}^2 + 2\theta\epsilon_t\epsilon_{t-1}].$$

$$= \underbrace{E[\epsilon_t^2]}_{\sigma^2} + \underbrace{E[\theta^2\epsilon_{t-1}^2]}_{\theta^2\sigma^2} + \underbrace{E[2\theta\epsilon_t\epsilon_{t-1}]}_{0}.$$

  Therefore,

$$Var(y_t) = E[y_t^2] = \sigma^2 + \theta^2\sigma^2 = \sigma^2(1 + \theta^2).$$

- Now consider the covariance.

$$Cov(y_t, y_{t-1}) = E[y_t y_{t-1}] = E[(\epsilon_t + \theta\epsilon_{t-1})(\epsilon_{t-1} + \theta\epsilon_{t-2})].$$

$$= E[\epsilon_t\epsilon_{t-1} + \theta^2\epsilon_{t-1}\epsilon_{t-2} + \theta\epsilon_{t-1}^2 + \theta\epsilon_t\epsilon_{t-2}].$$

$$= E[\theta\epsilon_{t-1}^2] = \theta\sigma^2.$$

  Note that if $s > 1$, $Cov(y_t, y_{t-s}) = 0$ because these terms have no common time periods. Thus $MA(1)$ is certainly ergodic because all the moments are time independent. In fact, $MA(p)$ is also ergodic by similar reasoning.

1.4 The Autoregressive Process

- Consider an $AR(1)$ process: $y_t = \phi y_{t-1} + \epsilon_t$ with ϵ_t iid. By backward substitution,

$$
\begin{aligned}
y_t &= \phi[\phi y_{t-2} + \epsilon_{t-1}] + \epsilon_t \\
&= \phi^2 y_{t-2} + \phi\epsilon_{t-1} + \epsilon_t \\
&= \phi^2[\phi y_{t-3} + \epsilon_{t-2}] + \phi\epsilon_{t-1} + \epsilon_t \\
&= \phi^3 y_{t-3} + \phi^2\epsilon_{t-2} + \phi\epsilon_{t-1} + \epsilon_t \\
&\vdots \\
&= \phi^s y_{t-s} + \phi^{s-1}\epsilon_{t-(s-1)} + \cdots + \phi\epsilon_{t-1} + \epsilon_t \\
&= \phi^s y_{t-s} + \sum_{j=0}^{s-1} \phi^j \epsilon_{t-j}.
\end{aligned}
$$

- Note that if $|\phi| < 1$, $\phi^s \to 0$ as $s \to \infty$.
- If $|\phi| > 1$, $\phi^s \to \infty$ as $s \to \infty$.
- If $|\phi| = 1$, $\phi^s \to 1$ as $s \to \infty$.
- So, if $|\phi| < 1$, then $y_t = \sum_{j=0}^{s-1} \phi^j \epsilon_{t-j}$. This is the condition for stationarity. [If $|\phi| \geq 1$, then y_t is non-stationary.]
- Assume stationarity, ie, $|\phi| < 1$. Thus,

$$E[y_t] = \sum_{j=0}^{\infty} \phi^j E[\epsilon_{t-j}] = 0.$$

Thus,

$$Var(y_t) = E[y_t^2] = E[(\sum_{j=0}^{s-1} \phi^j \epsilon_{t-j})(\sum_{j=0}^{s-1} \phi^j \epsilon_{t-j})].$$

Or,

$$E[y_t^2] = E[\epsilon_t + \phi\epsilon_{t-1} + \phi^2\epsilon_{t-2} + \ldots)(\epsilon_t + \phi\epsilon_{t-1} + \phi^2\epsilon_{t-2} + \ldots)].$$

Note that

$$E[\epsilon_t \epsilon_s] = 0 \;\forall\; s \neq t.$$

And,

$$E[\epsilon_t \epsilon_s] = \sigma^2 \;\forall\; s = t.$$

Thus,

$$E[y_t^2] = \sigma^2 + \phi^2\sigma^2 + \phi^4\sigma^2 \ldots$$

$$E[y_t^2] = \sigma^2(1 + \phi^2 + \phi^4 + \phi^6 \ldots) = \frac{\sigma^2}{1-\phi^2}.$$

1.5 More on the $AR(1)$ process

- Consider $y_t = \phi y_{t-1} + \epsilon_t$. Recall,

$$Var(y_t) = \frac{\sigma^2}{1-\phi^2}.$$

Thus, $Var(y_t) \to \infty$ as $\phi \to 1$. In other words, the process becomes more non-stationary as ϕ approaches 1.

- Recall that by backward substitution, we wrote,

$$y_t = \phi^s y_{t-s} + \phi^{s-1}\epsilon_{t-(s-1)} + \cdots + \phi\epsilon_{t-1} + \epsilon_t.$$

- Now consider the covariance:

$$Cov(y_t, y_{t-s}) = E[y_t y_{t-s}] = E[(\phi^s y_{t-s} + \phi^{s-1}\epsilon_{t-(s-1)} + \cdots + \phi\epsilon_{t-1} + \epsilon_t)y_{t-s}].$$

Note that, $E[\epsilon_t y_s] = 0$ if $s < t$ because $E[\epsilon_t y_s] = E[\epsilon_t \sum_{j=0}^{\infty} \phi^j \epsilon_{s-j}] = 0$. Then ensures,

$$E[y_t y_{t-s}] = E[\phi^s y_{t-s}^2]$$

since $E[\epsilon_{t-i} y_{t-s}] = 0 \; \forall \; i < s$. Thus,

$$E[\phi^s y_{t-s}^2] = \phi^s E[y_{t-s}^2].$$

$$= \phi^s Var[y_{t-s}].$$

$$= \phi^s Var[y_t],$$

by stationarity. So,

$$Cov(y_t, y_{t-s}) = E[y_t y_{t-s}] = \frac{\phi^s \sigma^2}{1-\phi^2}.$$

- Thus, y_t is ergodic.

1.6 Some definitions

- Autocovariance. (For any stochastic stationary process).

$$\gamma_s = Cov(y_t, y_{t-s}) = E[(y_t - \mu)(y_{t-s} - \mu)],$$

with $E[y_t] = \mu$. The sample equivalent is as follows:

$$r_s = \frac{1}{T-s}\sum_{t=s+1}^{T}(y_t - \bar{y})(y_{t-s} - \bar{y}).$$

- Autocorrelation.

$$\rho_s = \frac{Cov(y_t, y_{t-s})}{Var(y_t)} = \frac{\gamma_s}{\gamma_0}.$$

The sample equivalent is:

$$\frac{r_s}{r_0}.$$

- Correlogram: A picture of the autocorrelation.

Consider an $AR(1)$ process. Thus,

$$\gamma_0 = \frac{\sigma^2}{1-\phi^2},$$

and

$$\gamma_s = \phi^s \gamma_0.$$

Thus,

$$\rho_s = \frac{\gamma_s}{\gamma_0} = \phi^s.$$

See graph in notes.

Consider the $MA(1)$ process. $\gamma_0 = \sigma^2(1+\theta^2)$, $\gamma_1 = \theta\sigma^2$, $\gamma_2 = 0$, $\gamma_3 = 0$, $\ldots$. So

$$\rho_1 = \frac{\theta}{1+\theta^2}.$$

And $\rho_2 = 0$, $\rho_3 = 0$, $\ldots$. See graph in notes. Note that if the correlogram does not show $\rho_s \to 0$ eventually and is more or less a horizontal line, we might suspect non-stationarity.

- The Lag operator (L).

$$Ly_t = y_{t-1}.$$
$$L(Ly_t) = L^2 y_t = y_{t-2}.$$
$$L^s y_t = y_{t-s}.$$

We can also have leading:

$$L^{-1} y_t = y_{t+1}.$$
$$L^{-1}(L^{-1} y_t) = L^{-2} y_t = y_{t+2}.$$
$$L^{-s} y_t = y_{t+s}.$$

Consider the AR(1) process and we will now rearrange it using the L operator.

$$y_t = \phi y_{t-1} + \epsilon_t.$$

$$y_t = \phi L y_t + \epsilon_t.$$

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

Note that if $|x| < 1$, then $(1-x)^{-1} = 1 + x + x^2 + x^3 + \dots$ by a geometric progression that I should probably know by now. Thus,

$$y_t = (1 + \phi L + \phi^2 L^2 + \phi^3 L^3 + \dots)\epsilon_t.$$
$$y_t = (\epsilon_t + L\epsilon_t\phi + L^2\epsilon_t\phi^2 + L^3\epsilon_t\phi^3 + \dots).$$
$$y_t = (\epsilon_t + \epsilon_{t-1}\phi + \epsilon_{t-2}\phi^2 + \epsilon_{t-3}\phi^3 + \dots).$$
$$y_t = \sum_{j=0}^{\infty} \phi^j \epsilon_{t-j}.$$

Which we have previously shown to be true by backward substitution. Thus the L operator not only represents the lagging of an observation in time, but we can also apply practically all algebraic rules to it just as if it was a regular variable. Fascinating.

1.7 Estimation

- Model: $y = X\beta + \epsilon$ or in rows, $y_t = x_t'\beta + \epsilon_t$ with ϵ_t iid $(0, \sigma^2)$.
- Consider for a moment the relationship between X and ϵ.
  - Process Independent: Each element x_t is independent of ϵ_s $\forall\, t$ *and* s. Thus there is absolutely no relationship whatsoever. The $AR(1)$ process clearly does NOT have this property.
  - Contemporaneously Independent: Each element, x_t is independent of ϵ_t for $s \geq t$. But it can be dependent on ϵ_s for $s < t$. $AR(1)$ clearly DOES satisfy this property.

2 Week 2: 21 Jan - 25 Jan

2.1 A couple notes from last week

- When we were modeling $AR(1)$ processes for example, we said that shocks that occur in one period have an effect (diminishing) on all periods following it. This type of process is not uncommon in various economic models. Consider a simple consumption function:

  $$C_t = \phi y_{t-1} + \epsilon_{1t}.$$

  And the identity,

  $$y_t = C_t + i_t.$$

  Substituting,

  $$y_t - i_t = \phi y_{t-1} + \epsilon_{1t}.$$
  $$y_t = \phi y_{t-1} + \epsilon_{1t} + i_t.$$

  So as long as the propensity of consume is less than 1, we get a stationary process where shock in period t, affects all periods to follow. Granted this is a very simple model and most economic models have many more lags.

- The other item to mention here is the condition for stationarity in an $AR(p)$ process with p lags. Recall the simple model where $p = 1$:

  $$y_t = \phi y_{t-1} + \epsilon_t.$$

  And we determined that y_t is stationary iff $|\phi| < 1$. In an $AR(p)$ process,

  $$y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + \cdots + \phi_p y_{t-p} + \epsilon_t,$$

  the roots are the solutions to the equation,

  $$Z^p - \phi_1 Z^{p-1} - \phi_2 Z^{p-2} - \cdots - \phi_p = 0.$$

  This equation, in general, has p roots. The $AR(p)$ process is stationary if all roots, in absolute value, are less than 1. If the roots are complex (of the form $a + bi$), then $(a^2 + b^2)^{1/2}$ needs to be less than 1.

2.2 Types of independence for estimation procedures

- We have stated previously that in the simple regression equation,

$$y_t = x_t'\beta + \epsilon_t, \quad \epsilon_t \sim iid(0, \sigma^2)$$

the x's and the ϵ's can have three different relationships.

  - Process Independence: The x's and ϵ's are independent in all time periods.
  - Contemporaneously Independent: Current period independent.
  - Endogenous: x_t depends directly upon ϵ_t and thus all OLS procedures fail.

- If x is process independent of ϵ, then OLS is unbiased, consistent, and asymptotically normal. For example, the OLS coefficient, $\hat{\beta}$ is unbiased because,

$$E[\hat{\beta}] = \beta + E[(X'X)^{-1}X'\epsilon] = \beta + E[X'X)^{-1}X'] \underbrace{E[\epsilon]}_{0} = \beta.$$

- If x is contemporaneously independent of ϵ: OLS is biased. Consider the following example:

$$y_t = \phi_0 + \phi y_{t-1} + \epsilon_t, \quad \epsilon \sim iid(0, \sigma^2), \quad t = 1 \ldots T.$$

- If $\phi_0 \neq 0$, then,

$$\hat{\phi} = \frac{\sum (y_t - \bar{y})(y_{t-1} - \bar{y}_{-1})}{\sum (y_{t-1} - \bar{y}_{-1})^2}.$$

With $\bar{y} = \dfrac{1}{T} \sum_{t=1}^{T} y_t$ and $\bar{y}_{-1} = \dfrac{1}{T} \sum_{t=1}^{T} y_{t-1}$.

- If $\phi_0 = 0$, then,

$$\hat{\phi} = \frac{\sum (y_t y_{t-1})}{\sum y_{t-1}^2}.$$

- So why is $\hat{\phi}$ biased. Well, take the first example,

$$\hat{\phi} = \frac{\sum (y_t - \bar{y})(y_{t-1} - \bar{y}_{-1})}{\sum (y_{t-1} - \bar{y}_{-1})^2} = \frac{\sum (\phi_0 + \phi y_{t-1} + \epsilon_t - \phi_0 - \phi \bar{y}_{-1} - \bar{\epsilon})(y_{t-1} - \bar{y}_{-1})}{\sum (y_{t-1} - \bar{y}_{-1})^2}.$$

$$= \frac{\sum (\phi y_{t-1} + \epsilon_t - \phi \bar{y}_{-1} - \bar{\epsilon})(y_{t-1} - \bar{y}_{-1})}{\sum (y_{t-1} - \bar{y}_{-1})^2}.$$

$$= \frac{\phi \sum (y_{t-1} - \bar{y}_{-1}) \sum (\epsilon_t - \bar{\epsilon})(y_{t-1} - \bar{y}_{-1})}{\sum (y_{t-1} - \bar{y}_{-1})^2}.$$

$$= \phi + \frac{\sum(\epsilon_t - \bar{\epsilon})(y_{t-1} - \bar{y}_{-1})}{\sum(y_{t-1} - \bar{y}_{-1})^2}.$$

And note that $\bar{\epsilon} = \frac{1}{T}(\epsilon_1 + \epsilon_2 + \cdots + \epsilon_{t-1} + \epsilon_t + \cdots + \epsilon_T)$. And since $y_{t-1} = \phi_0 + \phi y_{t-2} + \epsilon_{t-1}$, y_{t-1} is correlated with ϵ_{t-1} so the expression on the right hand side of the last large equation will not be equal to 0. Because of the negative sign in front of the $\bar{\epsilon}$, the bias will be negative.

- To get an idea of just how much bias we are talking about. Consider the example above with $\phi_0 \neq 0$ and $\phi = 0.8$. Then,

$$\text{if } T = 25 \text{ then the bias } = -0.14.$$

$$\text{if } T = 50 \text{ then the bias } = -0.07.$$

$$\text{if } T = 150 \text{ then the bias } = -0.02.$$

- Consider the example above with $\phi_0 = 0$ and $\phi = 0.8$. Then, the situation is slightly better:

$$\text{if } T = 25 \text{ then the bias } = -0.05.$$

$$\text{if } T = 150 \text{ then the bias } = -0.01.$$

The intercept acts to "soak up" some of the bias.

2.3 Consistency under Contemporaneous Independence

- Given that the x's and the ϵ's are contemporaneously independent, can we show that they are consistent and asymptotically normal? Yes. Consider the model ($AR(1)$) above and let $\phi_0 = 0$. Thus,

$$\hat{\phi} = \frac{\sum(y_t y_{t-1})}{\sum y_{t-1}^2}.$$

Substituting in for y_t,

$$\hat{\phi} = \frac{\sum((\phi y_{t-1} + \epsilon_t) y_{t-1})}{\sum y_{t-1}^2}.$$

$$= \frac{\sum(\phi y_{t-1} y_{t-1} + \epsilon_t y_{t-1})}{\sum y_{t-1}^2}.$$

$$= \frac{\phi \sum y_{t-1}^2 + \sum(\epsilon_t y_{t-1})}{\sum y_{t-1}^2}.$$

$$= \phi + \frac{\sum(\epsilon_t y_{t-1})}{\sum y_{t-1}^2}.$$

Thus,

$$\hat{\phi} - \phi = \frac{\sum \epsilon_t y_{t-1}}{\sum y_{t-1}^2}.$$

- To show consistency, we need to find the probability limit of this last expression. Thus,

$$plim(\hat{\phi} - \phi) = plim\left(\frac{\sum \epsilon_t y_{t-1}}{\sum y_{t-1}^2}\right).$$
$$= \frac{plim\frac{1}{T}\sum \epsilon_t y_{t-1}}{plim\frac{1}{T}\sum y_{t-1}^2}.$$

  By Slutsky's Theorem.

- So, consider the denominator. Since this is a stationary ergodic sequence, by the ergodic theorem,

$$plim\frac{1}{T}\sum y_{t-1}^2 = E[y_{t-1}^2] = \frac{\sigma^2}{1-\phi^2}.$$

- So, consider the numerator. Since this is a stationary ergodic sequence, by the ergodic theorem,

$$plim\frac{1}{T}\sum y_{t-1}\epsilon_t = E[y_{t-1}\epsilon_t] = 0,$$

  because $\epsilon_t \sim iid$.

- Thus $plim(\hat{\phi} - \phi) = 0$ which implies consistency.

2.4 Mann and Wald Theorem

- Now, we know that $\epsilon_t \sim iid(0, \sigma^2)$, but to do inference, we would need normality. We get this by appealing to the central limit theorem (CLT).

- This is a CLT, which applies when x is not iid. Consider the regression equation,

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

- Assume:

  - 1) $\epsilon_t \sim iid(0, \sigma^2)$.
  - 2) $plim\frac{1}{T}\sum x_t x_t' = \Sigma_{xx}$ which is finite and positive definite. This assumption just states that the x's are sensible and not perfectly collinear.
  - $E[x_t\epsilon_t] = 0$. So we have no contemporary correlation.

- THEN:

$$plim \frac{1}{T} \sum x_t \epsilon_t = 0,$$

AND:

$$\sqrt{T}(\frac{1}{T} \sum x_t \epsilon_t) \longrightarrow^d N(0, \sigma^2 \Sigma_{xx}).$$

Craziness.

2.5 Inference Procedures

- Model:

$$y_t = \phi y_{t-1} + \epsilon_t, \quad \epsilon_t \sim iid(0, \sigma^2).$$

- Recall that in OLS, $\hat{\phi}$ is a consistent estimator of ϕ such that,

$$plim(\hat{\phi} - \phi) = plim \frac{1/T \sum y_{t-1} \epsilon_t}{1/T \sum y_{t-1}^2} \longrightarrow 0.$$

- Given our model, we can now apply the Mann-Wald theorem to find the limiting distribution of,

$$\frac{1}{T} \sum y_{t-1} \epsilon_t.$$

  - Note that condition 1 of Mann-Wald is satisfied because $\epsilon_t \sim iid(0, \sigma^2)$.
  - Note that condition 2 of Mann-Wald is satisfied because $plim \frac{1}{T} \sum y_{t-1}^2 = \frac{\sigma^2}{1 - \phi^2}$ as was shown previously. Thus, this matrix is positive definite and finite.
  - Note that condition 3 of Mann-Wald is satisfied because $E[y_{t-1} \epsilon_t] = 0$ because $E[\epsilon_t] = 0$.

Thus, the Mann-Wald theorem tells us:

$$\sqrt{T}\left(\frac{1}{T} \sum y_{t-1} \epsilon_t\right) \longrightarrow^d N(0, \sigma^2 [\frac{\sigma^2}{1 - \phi^2}]).$$

- Recall Cramer's Theorem: If $y_t \longrightarrow^d y \sim N(\mu, \sigma^2)$ and $plim(x_t) = a$, then $x_t y_t \longrightarrow^d ay \sim N(a\mu, a^2 \sigma^2)$.

- Consider again the expression we used above to prove consistency:

$$(\hat{\phi} - \phi) = \frac{1/T \sum y_{t-1} \epsilon_t}{1/T \sum y_{t-1}^2}.$$

Rewriting,

$$(\hat{\phi} - \phi) = (1/T \sum y_{t-1}^2)^{-1}(1/T \sum y_{t-1}\epsilon_t).$$

Multiplying both sides by $\sqrt{T}$,

$$\sqrt{T}(\hat{\phi} - \phi) = (1/T \sum y_{t-1}^2)^{-1}\sqrt{T}(1/T \sum y_{t-1}\epsilon_t).$$

We just showed that the second term on the right (via Mann-Wald) converges: $\sqrt{T}(1/T \sum y_{t-1}\epsilon_t) \longrightarrow^d N(0, \frac{\sigma^4}{1-\phi^2})$. Let this be our y_t in Cramer's theorem and let the first term on the right hand side be our value of a. In Cramer's Theorem, we will need the value of a^2 which is:

$$a^2 = plim[(1/T \sum y_{t-1}^2)^{-1}]^2 = plim(1/T \sum y_{t-1}^2)^{-2} = [\frac{\sigma^2}{1-\phi^2}]^{-2}.$$

Thus, via Cramer's theorem,

$$ay_t = (1/T \sum y_{t-1}^2)^{-1}\sqrt{T}(1/T \sum y_{t-1}\epsilon_t) \longrightarrow^d N\left(0, [\frac{\sigma^4}{1-\phi^2}][\frac{\sigma^2}{1-\phi^2}]^{-2}\right).$$

Since $ay_t = \sqrt{T}(\hat{\phi} - \phi)$,

$$\sqrt{T}(\hat{\phi} - \phi), \longrightarrow^d N(0, 1-\phi^2).$$

- Thus $Var(\sqrt{T}(\hat{\phi} - \phi))$ is approximately $1-\phi^2$ in large samples. And thus $Var(\hat{\phi} - \phi)$ is approximately $\frac{1}{T}(1-\phi^2)$ in large samples.

- How does this relate to the OLS variance? Recall:

$$Var_{OLS}(\hat{\phi} - \phi) = \frac{s^2}{\sum y_{t-1}^2}.$$

Where,

$$s^2 = \frac{1}{T-1}\underbrace{\sum(y_t - \hat{\phi}y_{t-1})}_{RSS}.$$

Now assuming,

$$plim(s^2) = \sigma^2,$$

then, multiplying both sides by T,

$$plim[T\ Var_{OLS}(\hat{\phi} - \phi)] = plim[\frac{s^2}{\frac{1}{T}\sum y_{t-1}^2}].$$

$$= \frac{plim(s^2)}{plim(\frac{1}{T}\sum y_{t-1}^2)}.$$

$$= \frac{\sigma^2}{\frac{\sigma^2}{1-\phi^2}} = 1 - \phi^2.$$

- So we get the result,

$$\frac{s^2}{\frac{1}{T}\sum y_{t-1}^2}$$

  is a consistent estimator for the variance of $\sqrt{T}(\hat{\phi} - \phi)$.

- So in sum, using OLS in a basic stationary time series model, such as AR(1) as shown here, asymptotically everything will be fine. No worries. Of course this can easily be extended to more complicated models with more lags which thankfully we won't be doing.

2.6 Maximum Likelihood Estimation

- As we did in regular OLS regression, having MLE back up our results proves that they are robust. We will now show that the results just obtained for OLS is shown to be true using MLE in the time series type of model.

- If we have unknown parameters, ψ, the ML esimates, $\hat{\psi}$, are consistent and,

$$\sqrt{T}(\hat{\psi} - \psi) \longrightarrow^d N\Big(0, (lim(\frac{1}{T}I(\psi))^{-1}\Big)$$

  ONLY IF THE PROCESSES ARE ERGODIC.

- Recall the information matrix,

$$I(\psi) = E\Big[\frac{-\partial^2 logL(\psi)}{\partial\psi\partial\psi'}\Big].$$

- To construct the likelihood, L, we make use of the fact that in terms of joint probability distributions:

$$f(y_3, y_2, y_1) = f(y_3|y_2, y_1) \cdot f(y_2, y_1) = f(y_3|y_2, y_1) \cdot f(y_2|y_1) \cdot f(y_1).$$

  And we could extend this for any number of y's using conditional probabilities.

- In general, if we have T observations,

$$L(\psi) = f(y_T|y_{T-1}, y_{T-2}, y_{T-3}, \ldots, y_1) \cdot f(y_{T-1}|y_{T-2}, y_{T-3}, y_{T-4}, \ldots, y_1) \cdot \cdots \cdot f(y_2|y_1) \cdot f(y_1).$$

  Or in much simplier terms,

$$L(\psi) = \left[\prod_{t=2}^{T} f(y_t | \underbrace{y_{t-1}, y_{t-2}, \ldots, y_1)}_{Info\ Set\ at\ time\ t-1} \right] f(y_1).$$

- This is useful because time series models are usually written in conditional firm, eg:

  $$AR(1) : y_t = \phi y_{t-1} + \epsilon_t, \quad \epsilon_t \sim iidN(0, \sigma^2), \quad |\phi| < 1.$$

  This implies,

  $$y_t | y_{t-1} \sim N(\phi y_{t-1}, \sigma^2).$$

  Thus if you know y_{t-1}, you know more about y_t then if you don't know y_{t-1}.

- Continuing with the analysis,

  $$f(y_t | y_{t-1}) = \frac{1}{(2\pi\sigma^2)^{1/2}} exp\left[\frac{-1}{2\sigma^2}(y_t - \phi y_{t-1})^2 \right].$$

  Taking logs,

  $$log(f(y_t|y_{t-1})) = log(2\pi\sigma^2)^{-1/2} log\left(exp\left[\frac{-1}{2\sigma^2}(y_t - \phi y_{t-1} \right))^2 \right]).$$

  $$log(f(y_t|y_{t-1})) = -1/2 log(2\pi) - 1/2 log(\sigma^2) + \left[\frac{-1}{2\sigma^2}(y_t - \phi y_{t-1})^2 \right].$$

  And when taking the product of all of this from $t = 2$ to T, we get,

  $$log\ L(\phi, \sigma^2) = -\frac{T-1}{2} log(2\pi) - \frac{T-1}{2} log(\sigma^2) - \frac{1}{2\sigma^2} \sum_{t=2}^{T} (y_t - \phi y_{t-1})^2 + log\ f(y_1).$$

  And that's our log likelihood function.

3 Week 3: 28 Jan - 1 Feb

3.1 More on Maximum Likelihood: Time Series

- Recall, we derived the log likelihood function for a simple $AR(1)$ process last week:

$$log\ L(\phi, \sigma^2) = -\frac{T-1}{2} log(2\pi) - \frac{T-1}{2} log(\sigma^2) - \frac{1}{2\sigma^2} \sum_{t=2}^{T} (y_t - \phi y_{t-1})^2 + log\ f(y_1).$$

- We could assume that the last term $f(y_1)$ is either fixed or it might have some distribution associated with it. If we assume it is constant, then everything becomes easier because when we take derivatives, it drops out. (This also makes sense if we consider y_1 the "first" observation in time so it would be a constant.) Assuming that it follows some distribution actually doesn't introduce too many problems. Since MLE is always done in the background of asymptotic results, the probability distribution of any single observation, ie $f(y_1)$, has zero influence in the limit. Thus for simplicity, assume $f(y_1)$ is constant.

- Now maximize the log likelihood function with respect to ϕ,

$$\frac{\partial logL}{\partial \phi} \Rightarrow -\frac{1}{2\sigma^2} \sum_{t=2}^{T} -2y_{t-1}(y_t - \phi y_{t-1}) = 0.$$

Solving for $\hat{\phi}$,

$$-\frac{1}{2\sigma^2} \sum_{t=2}^{T} -2y_{t-1}(y_t - \phi y_{t-1}) = 0.$$

$$\sum_{t=2}^{T} y_{t-1}(y_t - \phi y_{t-1}) = 0.$$

$$\sum_{t=2}^{T} y_{t-1}y_t - \phi y_{t-1}^2 = 0.$$

$$\sum_{t=2}^{T} y_{t-1}y_t - \sum_{t=2}^{T} \phi y_{t-1}^2 = 0.$$

$$\sum_{t=2}^{T} y_{t-1}y_t = \phi \sum_{t=2}^{T} y_{t-1}^2.$$

$$\hat{\phi} = \frac{\sum_{t=2}^{T} y_{t-1}y_t}{\sum_{t=2}^{T} y_{t-1}^2}$$

Note that this is equivalent to the OLS estimator of $\hat{\phi}$!! Thus, we have robustness.

- Now maximize the log likelihood function with respect to σ^2,

$$\frac{\partial logL}{\partial \sigma^2} \Rightarrow \frac{-(T-1)}{2\sigma^2} + \frac{1}{2(\sigma^2)^2}\sum_{t=2}^{T}(y_t - \phi y_{t-1})^2 = 0.$$

Solving for $\hat{\sigma^2}$,

$$\frac{-(T-1)}{2\sigma^2} + \frac{1}{2(\sigma^2)^2}\sum_{t=2}^{T}(y_t - \phi y_{t-1})^2 = 0.$$

$$\frac{(T-1)}{2\sigma^2} = \frac{1}{2(\sigma^2)^2}\sum_{t=2}^{T}(y_t - \phi y_{t-1})^2.$$

$$\frac{(T-1)\sigma^2}{2} = \frac{1}{2}\sum_{t=2}^{T}(y_t - \phi y_{t-1})^2.$$

$$(T-1)\sigma^2 = \sum_{t=2}^{T}(y_t - \phi y_{t-1})^2.$$

$$\hat{\sigma^2} = \frac{1}{T-1}\sum_{t=2}^{T}(y_t - \phi y_{t-1})^2.$$

$$\hat{\sigma^2} = \frac{1}{T-1}\sum_{t=2}^{T}\hat{\epsilon}_t^2.$$

Note that $\hat{\sigma}^2_{OLS} = \frac{1}{T-1}\sum_{t=1}^{T}\hat{\epsilon}_t^2$. So in the OLS estimate of the variance, we divide the RSS by one less than the number of observations. In MLE, we just divide by the exact number of observations. (Note the limits on the summation). This is because again MLE is always done assuming T is very large so the bias in the ML estimator goes to zero, as T goes to infinity. So, in the limit, the OLS and ML estimators for the variance are the same. Again, robustness.

- Now, to do any sort of inference procedure on $\hat{\phi}$ or on $\hat{\sigma}^2$, we would need their standard errors (ie, variances). Thus, we would like to compute a variance covariance matrix for these two coefficients.

- In ML,

$$Var\begin{pmatrix}\hat{\phi}\\ \hat{\sigma}^2\end{pmatrix}, \tag{1}$$

can be approximated by,

$$-\begin{bmatrix}\frac{\partial^2 logL}{\partial\hat{\phi}^2} & \frac{\partial^2 logL}{\partial\hat{\phi}\partial\hat{\sigma}^2}\\ \frac{\partial^2 logL}{\partial\hat{\phi}\partial\hat{\sigma}^2} & \frac{\partial^2 logL}{\partial(\hat{\sigma}^2)^2}\end{bmatrix}^{-1}. \tag{2}$$

- So, now to work out all those (3) second order conditions and evalute them at $(\hat{\phi}, \hat{\sigma}^2)$:
  Upper left:

$$\frac{\partial^2 logL}{\partial \hat{\phi}^2} = -\frac{1}{\hat{\sigma}^2}\sum_{t=2}^{T} y_{t-1}^2.$$

  Upper right and lower left:

$$\frac{\partial^2 logL}{\partial \hat{\phi} \partial \hat{\sigma}^2} = -\frac{1}{(\hat{\sigma}^2)^2}\sum_{t=2}^{T} y_{t-1}(y_t - \hat{\phi} y_{t-1}) = 0.$$

  (After substituting in $\hat{\phi}$ which zeros the summation.)
  Lower right:

$$\frac{\partial^2 logL}{\partial (\hat{\sigma}^2)^2} = \frac{T-1}{2(\hat{\sigma}^2)^2} - \frac{1}{(\hat{\sigma}^2)^3}\underbrace{\sum_{t=2}^{T}(y_t - \phi y_{t-1})^2}_{(T-1)\hat{\sigma}^2}.$$

$$= \frac{T-1}{2(\hat{\sigma}^2)^2} - \frac{1}{(\hat{\sigma}^2)^3}(T-1)\hat{\sigma}^2.$$

$$= \frac{T-1}{2(\hat{\sigma}^2)^2} - \frac{T-1}{(\hat{\sigma}^2)^2}.$$

$$= -\frac{T-1}{2(\hat{\sigma}^2)^2}.$$

- So substituting into our matrix above with the second derivatives that we just derived:

$$Var\begin{pmatrix} \hat{\phi} \\ \hat{\sigma}^2 \end{pmatrix} = -\begin{bmatrix} \frac{\partial^2 logL}{\partial \hat{\phi}^2} & \frac{\partial^2 logL}{\partial \hat{\phi} \partial \hat{\sigma}^2} \\ \frac{\partial^2 logL}{\partial \hat{\phi} \partial \hat{\sigma}^2} & \frac{\partial^2 logL}{\partial (\hat{\sigma}^2)^2} \end{bmatrix}^{-1} = -\begin{bmatrix} -\frac{1}{\hat{\sigma}^2}\sum_{t=2}^{T} y_{t-1}^2 & 0 \\ 0 & -\frac{T-1}{2(\hat{\sigma}^2)^2} \end{bmatrix}^{-1}. \quad (3)$$

$$= \begin{bmatrix} \frac{1}{\hat{\sigma}^2}\sum_{t=2}^{T} y_{t-1}^2 & 0 \\ 0 & \frac{T-1}{2(\hat{\sigma}^2)^2} \end{bmatrix}^{-1}. \quad (4)$$

$$= \begin{bmatrix} \frac{\hat{\sigma}^2}{\sum_{t=2}^{T} y_{t-1}^2} & 0 \\ 0 & \frac{2(\hat{\sigma}^2)^2}{T-1} \end{bmatrix}. \quad (5)$$

- In sum: We have under OLS,

$$\hat{\phi} = \frac{\sum_{t=1}^{T} y_{t-1}y_t}{\sum_{t=1}^{T} y_{t-1}^2} \text{ and } Var(\hat{\phi}) = \frac{s^2}{\sum_{t=1}^{T} y_{t-1}^2}.$$

- Under ML,

$$\hat{\phi} = \frac{\sum_{t=1}^{T} y_{t-1}y_t}{\sum_{t=1}^{T} y_{t-1}^2} \text{ and } Var(\hat{\phi}) = \frac{\hat{\sigma}^2}{\sum_{t=2}^{T} y_{t-1}^2}.$$

- Where,

$$s^2 = \frac{1}{T-1}\sum_{t=1}^{T}\hat{\epsilon}_t^2,$$

and,

$$\hat{\sigma}^2 = \frac{1}{T-1}\sum_{t=2}^{T}\hat{\epsilon}_t^2.$$

Note the different ranges of the summations.

- So if ML yields exactly the same estimators when we look at large T, what's the use? Well, MLE is good for non-linear estimation as will be shown presently.

3.2 Non-Linear Estimation - Least Squares and MLE

- Consider the following nonlinear model:

$$y_t = g(x_t, \beta) + \epsilon_t, \quad \epsilon_t \sim iid\ N(0, \sigma^2).$$

- Assume that x_t is independent of ϵ_t. Note that $y_t|x_t \sim N(g(x_t, \beta), \sigma^2)$.

- Thus the log likelihood function would take the form:

$$log\ L(\beta, \sigma^2) = -\frac{T}{2}log(2\pi) - \frac{T}{2}log(\sigma^2) - \frac{1}{2\sigma^2}\sum_{t=1}^{T}[y_t - g(x_t, \beta)]^2.$$

Or in other terms,

$$log\ L(\beta, \sigma^2) = -\frac{T}{2}log(2\pi) - \frac{T}{2}log(\sigma^2) - \frac{1}{2\sigma^2}\sum_{t=1}^{T}\epsilon_t^2(\beta).$$

Where, $\epsilon_t = y_t - g(x_t, \beta)$.

- So maximizing the log likelihood function with respect to β is the same as minimizing the last term because it is the only term involving a β. So FOC for ML:

$$\frac{\partial logL}{\partial \beta} \Rightarrow -\frac{1}{2\sigma^2}\sum_{1}^{T} 2\epsilon_t(\beta)\frac{\partial \epsilon_t(\beta)}{\partial \beta} = 0.$$

- Define a new variable, z_t as follows:

$$z_t = -\frac{\partial \epsilon_t(\beta)}{\partial \beta} = \frac{\partial g(x_t, \beta)}{\partial \beta}.$$

- Note that Z_t depends only on x_t and β and is independent of ϵ_t because x_t is assumed to be independent of ϵ_t.

- We'll continue the next step in the next section.

3.3 More on the Nonlinear ML Estimation

- Recall the model:

$$y_t = g(x_t, \beta) + \epsilon_t \quad \epsilon_t \sim iid\ N(0, \sigma^2).$$

- With log-likelihood function:

$$log\ L(\beta, \sigma^2) = -\frac{T}{2}log(2\pi) - \frac{T}{2}log(\sigma^2) - \frac{1}{2\sigma^2}\sum_{t=1}^{T}\epsilon_t^2(\beta).$$

  Where, $\epsilon_t = y_t - g(x_t, \beta)$.

- And $FOC(\beta)$:

$$\frac{\partial logL}{\partial \beta} \Rightarrow -\frac{1}{2\hat{\sigma}^2}\sum_{1}^{T} 2\epsilon_t(\beta)\frac{\partial \epsilon_t(\beta)}{\partial \beta} = 0.$$

- Define a new variable, z_t as follows:

$$z_t = -\frac{\partial \epsilon_t(\beta)}{\partial \beta} = \frac{\partial g(x_t, \beta)}{\partial \beta}.$$

- And $FOC(\sigma^2)$,

$$\frac{\partial logL}{\partial \sigma^2} \Rightarrow -\frac{T}{2\hat{\sigma}^2} + \frac{1}{2(\hat{\sigma}^2)^2}\sum_{1}^{T}\epsilon_t^2(\beta) = 0.$$

Solving for $\hat{\sigma}^2$,

$$\frac{T}{2\hat{\sigma}^2} = \frac{1}{2(\hat{\sigma}^2)^2} \sum_1^T \epsilon_t^2(\beta).$$

$$\hat{\sigma}^2 = \frac{1}{T} \sum_1^T \epsilon_t^2(\hat{\beta}).$$

- From this, we know that $\hat{\beta}$ solves $\sum_1^T 2\epsilon_t(\beta)\dfrac{\partial \epsilon_t(\beta)}{\partial \beta} = 0$. So what is the variance of $\hat{\beta}$? We will use the fact that:

$$Var(\hat{\psi}) \simeq \underbrace{I(\hat{\psi})}_{Information\ Matrix} = -E\left[\frac{\partial^2 log\ L}{\partial \psi \partial \psi'}\right]^{-1}.$$

- So, we'll now take second derivatives of the log likelihood function with respect to β and σ^2. The top left element of this matrix is:

$$-E\left[\frac{\partial^2 log\ L}{\partial \beta \partial \beta'}\right].$$

$$= -\Bigg[-\frac{1}{\sigma^2}\sum_{t=1}^T E\Big[\frac{\partial \epsilon_t(\beta)}{\partial \beta}\frac{\partial \epsilon_t(\beta)}{\partial \beta'}\Big] - \underbrace{\frac{1}{\sigma^2}\sum_{t=1}^T E\Big[\epsilon_t(\beta)\frac{\partial^2 \epsilon_t(\beta)}{\partial \beta \partial \beta'}\Big]}_{0}\Bigg],$$

because x_t and ϵ_t are independent. Thus,

$$\frac{1}{\sigma^2}\sum_{t=1}^T E\Big[\frac{\partial \epsilon_t(\beta)}{\partial \beta}\frac{\partial \epsilon_t(\beta)}{\partial \beta'}\Big] = \frac{1}{\sigma^2}\sum_{t=1}^T E\Big[z_t z_t'\Big].$$

- We can also find the cross 2nd order derivatives:

$$-E\left[\frac{\partial^2 log\ L}{\partial \beta \partial \sigma^2}\right] = -\frac{1}{(\sigma^2)^2}\sum_{t=1}^T \underbrace{E[\epsilon_t]}_{0} E[\frac{\partial \epsilon_t}{\partial \beta}] = 0.$$

- And finally, we'll get the variance of $\hat{\sigma}^2$ from,

$$-E\left[\frac{\partial^2 log\ L}{\partial (\sigma^2)^2}\right] = -\frac{T}{2(\sigma^2)^2} + \frac{2}{2(\sigma^2)^3}\sum_{t=1}^T \underbrace{E[\epsilon_t^2(\beta)]}_{T\sigma^2} = \frac{T}{2(\sigma^2)^2}.$$

- So, now we can complete the variance/covariance matrix:

$$Var\begin{pmatrix} \hat{\beta} \\ \hat{\sigma}^2 \end{pmatrix} \simeq - \begin{bmatrix} \dfrac{\partial^2 logL}{\partial \hat{\beta}^2} & \dfrac{\partial^2 logL}{\partial \hat{\beta} \partial \hat{\sigma}^2} \\ \dfrac{\partial^2 logL}{\partial \hat{\beta} \partial \hat{\sigma}^2} & \dfrac{\partial^2 logL}{\partial (\hat{\sigma}^2)^2} \end{bmatrix}^{-1} = \begin{bmatrix} \dfrac{1}{\sigma^2} \sum E[z_t z_t'] & 0 \\ 0 & \dfrac{T}{2(\sigma^2)^2} \end{bmatrix}^{-1}. \quad (6)$$

$$= \begin{bmatrix} \dfrac{\sigma^2}{\sum E[z_t z_t']} & 0 \\ 0 & \dfrac{2(\sigma^2)^2}{T} \end{bmatrix}. \quad (7)$$

- Note that in this derivation, we have to take expectations, where in the last var/cov matrix we derived there were none. This is due to the random variable, z_t, included in the analysis.

- Thus, we can approximate using sample moments, $Var(\hat{\beta}) = \hat{\sigma}^2 \Big[\sum_{t=1}^{T} z_t z_t' \Big]^{-1}$.

- In summary, to estimate the non-linear model, minimize $\sum \epsilon_t^2 \iff$ minimize $\sum (y_t - g(x_t, \beta))^2$ with respect to β. Note that if the model is linear, then $z_t = X$. Thus we have our standard result:

$$Var(\hat{\beta}) = \hat{\sigma}^2 \Big[X'X \Big]^{-1}.$$

3.4 MLE: Moving Average Model

- Consider the model:

$$y_t = \epsilon_t + \theta \epsilon_{t-1}, \quad \epsilon_t \; iid \; N(0, \sigma^2).$$

- We would like to find an estimate for θ. First define $\epsilon(\theta)$ given that $\epsilon_0 = 0$. So we assume that the very first observation takes on its expected value. If $\epsilon_0 = 0$, then

$$\epsilon_1 = y_1 - \theta \epsilon_0 = y_1.$$
$$\epsilon_2 = y_2 - \theta \epsilon_1 = y_2 - \theta y_1.$$
$$\epsilon_3 = y_3 - \theta \epsilon_2 = y_3 - \theta y_2 + \theta^2 y_1.$$
$$\vdots$$
$$\epsilon_t = y_t - \theta y_{t-1} + \theta^2 y_{t-2} - \theta^3 y_{t-3} + \cdots + (-\theta)^{t-1} y_1.$$

- So $f(y_t | I_{t-1}) \sim N(\theta \epsilon_{t-1}, \sigma^2)$.

- Thus, following the usual steps, we have the log-likelihood function:

$$log \; L(\theta, \sigma^2) = -\frac{T}{2} log(2\pi) - \frac{T}{2} log\sigma^2 - \frac{1}{2\sigma^2} \sum_{t=1}^{T} (y_t - \theta \epsilon_{t-1})^2.$$

$$= -\frac{T}{2}log(2\pi) - \frac{T}{2}log\sigma^2 - \frac{1}{2\sigma^2}\sum_{t=1}^{T}\epsilon_t^2(\theta).$$

- The log L function is maximized if $\sum \epsilon_t^2(\theta)$ is minimized. ie, $\hat{\theta}$ solves:

$$min_\theta \sum_{t=1}^{T}\epsilon_t^2(\theta).$$

- The $Var(\hat{\theta})$ is therefore:

$$Var(\hat{\theta}) = \frac{\hat{\sigma}^2}{\sum(z_t(\theta))^2},$$

where,

$$z_t = \frac{\partial\epsilon_t(\theta)}{\partial\theta}.$$

3.5 Review of Asymptotic Theory

- In most models, we cannot get unbiased estimators, but we can find consistent estimators. Hence the need for asymptotics.
- An estimator is unbiased if you take many samples and calculate means and the distribution of the means is centered at the population mean.
- An estimator is consistent if you take one sample and keep on adding observations on to it and if you have added enough observations, it is very likely that the estimator is the best estimate of the true parameter.
- More on this next week.

4 Week 4: 4 Feb - 8 Feb

4.1 Asymptotic Theory

- Assume z_i is a random sample that is *iid* (μ, σ^2).
- Let the sample size $= n$. We would like to find an estimate for μ.
- Let x_n be the following:

$$x_n = \frac{1}{n} \sum_{i=1}^{n} z_i.$$

Of course, x_n, is just the sample mean. Since all the z_i's come from the distribution above, Then,

$$E[x_n] = \mu.$$

And,

$$Var(x_n) = var\left(\frac{1}{n} \sum_{i=1}^{n} z_i\right) = \frac{1}{n^2} Var\left(\sum_{i=1}^{n} z_i\right) = \frac{1}{n^2} \sum_{i=1}^{n} \Bigg(Var(z_i)\Bigg) = \frac{1}{n^2}(n \cdot \sigma^2) = \frac{\sigma^2}{n}.$$

- If we took many samples of size n and drew the frequency distribution of x_n, we would find that the distribution is centered around μ. But note the variance. As $n \longrightarrow \infty$, $Var(x_n) \longrightarrow 0$. So as the sample sizes gets large, the variance of the sample mean goes to zero. In other words, the distribution of x_n collapses to a spike at the true mean, μ. Thus we say:

$$plim_{n \to \infty}(x_n) = \mu.$$

- We'll get to the definition of *plim* shortly, but first note another interesting result regarding the variance of x_n. It is independent of the population size. It is only a function of the sample size. Therefore, when using a sample to estimate a parameter, you don't have to use a huge sample if the population is larger. The same size sample will be equally efficient. For instance, one could estimate a parameter regarding the population of the UK with the same size sample as one would use to estimate a parameter regarding the US population. Interesting but not very well known in practice.
- Definition: PROBABILITY LIMIT. The probability limit (or *plim*) of a sequence x_n is equal to a, iff:

$$plim_{n \to \infty} Prob\Big((|x_n - a|) < \epsilon\Big) = 1 \; \forall \; \epsilon > 0.$$

- Note that the proceeding definition is the actual formal definition of a *plim*, but in practice, we use a slightly different definition. The $plim(x_n) = a$ iff:

$$\underbrace{\lim_{n\to\infty} E[x_n] = a}_{\textit{Spike at a}} \text{ AND } \underbrace{\lim_{n\to\infty} Var(x_n) = 0}_{\textit{Spike width equals zero}}.$$

- Two additional points should be made about these definitions.
  - 1) x_n can be a sequence, a vector, or a matrix.
  - 2) $plim(x_n) = a \nRightarrow E[x_n] = a$. Basically this says that consistancy does not imply unbiasedness. For instance consider the OLS and ML estimators for σ^2. The OLS estimator is unbiased but the ML estimator is consistant and only unbiased in the limit.

- Finding unbiased estimators is sometimes difficult but usually you can find a consistent estimator.

4.1.1 Useful Theorems in Asymptotics

- 1) The Weak Law of Large Numbers. (Khintchine) If $z_i, i = 1, 2, \ldots$ are *iid* with mean, μ, then

$$plim \frac{1}{n} \sum_{i=1}^{n} z_i = \mu.$$

  Note here that we have shown by the ergodic theorem that you don't actually need *iid*, but rather just ergodicity. Hence this theorem is rather "weak."

- 2) Convergence in Distribution. Suppose that you have a sequence of random variables, x_n with cummulative density function (cdf), F_n. (Recall $F_n(x_0) = Prob(x_n \leq x_0)$). Consider another random variable, x, such that F is the cdf of x. Suppose:

$$\lim_{n\to\infty} F_n(x_0) = F(x_0)\ \forall\ x_0 \text{ where } F \text{ is continuous.}$$

  Then: x_n converges in distribution to x.

- 3) The Central Limit Theorem. Consider z_i *iid* with $i = 1, 2, \ldots$. Recall from above that $Var(\frac{1}{n}\sum_{i=1}^{n} z_i) = \frac{\sigma^2}{n}$. This implies:

$$Var\left[\sqrt{n}(\frac{1}{n}\sum_{i=1}^{n} z_i)\right] = \sigma^2.$$

  Note that as $n \longrightarrow \infty$, this variance is constant. The $\sqrt{n}$ acts as a buffer from it collapsing to zero as before. Thus, by the CLT,

$$\sqrt{n}(\frac{1}{n}\sum_{i=1}^{n} z_i - \mu) \longrightarrow^D N(0, \sigma^2).$$

So if we have a consistent estimate of σ^2, call is s^2, then for large n,

$$\sqrt{n}(\frac{1}{n}\sum_{i=1}^{n} z_i - \mu) \sim_{approx} N(0, s^2).$$

And,

$$(\frac{1}{n}\sum_{i=1}^{n} z_i) \sim_{approx} N(\mu, \frac{s^2}{n}).$$

Note, this also works for z_i, random *iid* vectors with mean, μ and variance, Ω. Then,

$$\sqrt{n}(\frac{1}{n}\sum_{i=1}^{n} z_i - \mu) \sim_{approx} N(0, \Omega).$$

- 4) Slutsky's Theorem. If $plim(x_n) = a$ then for a continuous function g, $plim(g(x_n)) = g(a)$. If $plim(x_n) = a$ and $plim(y_n) = b$, then $plim(x_n y_n) = ab$ and $plim(\frac{x_n}{y_n}) = \frac{a}{b}$ if $b \neq 0$.
- 5) Cramer's Theorem. Suppose y_n is such that $y_n \rightarrow^D y$ and x_n is such that $plim$ $(x_n) = a$. Then:
  - a) $y_n + x_n \rightarrow^D y + a$.
  - b) $x_n y_n \rightarrow^D ay$.
  - c) $\frac{y_n}{x_n} \rightarrow^D \frac{y}{a}$.

  (eg) Suppose $y \sim N(\mu, \sigma^2)$. Then $x_n y_n \rightarrow ay \sim N(a\mu, a^2\sigma^2)$. We can also do this with vectors. Suppose $y_n \sim N(\mu, \Sigma)$ and $plim(X_n) = X$. Then,

$$X_n y_n \rightarrow^D Xy \sim N(X\mu, X\Sigma X').$$

4.2 Asymptotic Distribution of OLS

- Consider the following model:

$$y = X\beta + \epsilon, \quad \epsilon \sim iid(0, \sigma^2), \quad i = 1 \ldots n.$$

- Or in another form: $y = x_i'\beta + \epsilon_i$ with x_i' as the i^{th} row of X.

- Assume the x_i are random *iid* regressors with $Var(x_i) = E[x_i x_i'] = \Sigma_{xx}$, or otherwise known as the "moment matrix."
- Assume x_i and ϵ_i are independent.
- We would first like to check that the $\hat{\beta}$ estimator is consistant. Consider the definition of $\hat{\beta}$:

$$\hat{\beta} = (X'X)^{-1}X'y.$$

Substituting in y,

$$\hat{\beta} = (X'X)^{-1}X'[X\beta + \epsilon] = \beta + (X'X)^{-1}X'\epsilon.$$

Thus,

$$\hat{\beta} - \beta = (X'X)^{-1}X'\epsilon.$$

Or substituing in, (just algebra):

$$\hat{\beta} - \beta = \Big[\sum_{i=1}^{n} x_i x_i'\Big]^{-1}\Big[\sum_{i=1}^{n} x_i \epsilon_i\Big].$$

Multiplying both terms on the right by 1/n (because of the inverse they cancel themselves out):

$$\hat{\beta} - \beta = \Big[\frac{1}{n}\sum_{i=1}^{n} x_i x_i'\Big]^{-1}\Big[\frac{1}{n}\sum_{i=1}^{n} x_i \epsilon_i\Big].$$

Taking the probability limit:

$$plim(\hat{\beta} - \beta) = plim\Bigg[\Big[\frac{1}{n}\sum_{i=1}^{n} x_i x_i'\Big]^{-1}\Big[\frac{1}{n}\sum_{i=1}^{n} x_i \epsilon_i\Big]\Bigg].$$

By Slutsky's Theorem:

$$plim(\hat{\beta} - \beta) = \Big[plim\frac{1}{n}\sum_{i=1}^{n} x_i x_i'\Big]^{-1}\Big[plim\frac{1}{n}\sum_{i=1}^{n} x_i \epsilon_i\Big].$$

Since $x_i x_i'$ is an *iid* sequence, and $E[x_i x_i'] = \Sigma_{xx}$, then by the Weak Law of Large Numbers (WLLN),

$$plim\frac{1}{n}\sum_{i=1}^{n} x_i x_i' = \Sigma_{xx}.$$

Also, $x_i\epsilon_i$ is an *iid* sequence with $E[x_i\epsilon_i] = 0$ because of independence. Thus, by the WLLN,

$$plim\frac{1}{n}\sum_{i=1}^{n} x_i\epsilon_i = 0.$$

Thus, substituting these last two equations in,

$$plim(\hat{\beta} - \beta) = \Big[\Sigma_{xx}\Big]^{-1}\Big[0\Big] = 0.$$

Thus, $\hat{\beta}$ is a consistant estimator.

4.2.1 Asymptotic Distribution

- Consider again the sequence $x_i\epsilon_i$ which is *iid* and due to independence, $E[x_i\epsilon_i] = 0$. The variance of the sequence is as follows (noting that ϵ_i is a scaler):

$$Var(x_i\epsilon_i) = E[x_i\epsilon_i(x_i\epsilon_i)'] = E[x_i\epsilon_i\epsilon_i'x_i'] = E[x_i\epsilon_i^2x_i'] = E[\epsilon_i^2x_ix_i'] = E[\epsilon_i^2]E[x_ix_i'] = \sigma^2\Sigma_{xx}.$$

- Thus we get a key result by the Central Limit Theorem:

$$\sqrt{n}\Big(\frac{1}{n}\sum_{i=1}^{n} x_i\epsilon_i\Big) \longrightarrow^d N(0, \sigma^2\Sigma_{xx}).$$

But we want to find the limiting distribution of:

$$\hat{\beta} - \beta = \Big[\frac{1}{n}\sum_{i=1}^{n} x_ix_i'\Big]^{-1}\underbrace{\Big[\frac{1}{n}\sum_{i=1}^{n} x_i\epsilon_i\Big]}_{We're\ golden}.$$

- So, the second term we have just shown that it converges to a normal distribution with mean 0 and variance $\sigma_2\Sigma_{xx}$. Now for the first term. Via Cramer's Theorem:

$$\sqrt{n}(\hat{\beta} - \beta) = \sqrt{n}\Big[\frac{1}{n}\sum_{i=1}^{n} x_ix_i'\Big]^{-1}\Big[\frac{1}{n}\sum_{i=1}^{n} x_i\epsilon_i\Big] \longrightarrow^d N(0, \Sigma_{xx}^{-1}\sigma^2\Sigma_{xx}\Sigma_{xx}^{-1}).$$

Why all this mess? Recall cramer's theorem: Suppose $y_n \sim N(\mu, \Sigma)$ [y_n is our second term] and $plim(X_n) = X$. [X_n is our first term] Then, $X_ny_n \rightarrow^D Xy \sim N(X\mu, X\Sigma X')$. [Notice that the variance equals the variance of the second term pre-multiplied by the plim of the first term and post-multiplied by the inverse of the plim of the first term. In our case, $plim(X_n) = \Sigma_{xx}^{-1} = \Sigma_{xx}^{-1'}$ because it is symmetric.] Thus,

$$\sqrt{n}(\hat{\beta}-\beta)=\sqrt{n}\Big[\frac{1}{n}\sum_{i=1}^{n}x_i x_i'\Big]^{-1}\Big[\frac{1}{n}\sum_{i=1}^{n}x_i\epsilon_i\Big]\longrightarrow^d N(0,\sigma^2\Sigma_{xx}^{-1}).$$

- Now σ^2 can be approximated by s^2 and Σ_{xx} can be approximated by $\frac{1}{n}(X'X)$. Thus,

$$\sqrt{n}(\hat{\beta}-\beta)\longrightarrow^d N(0,s^2[\frac{1}{n}(X'X)]^{-1}).$$

  Factoring out the $\sqrt{n}$,

$$\hat{\beta}-\beta\longrightarrow^d N(0,s^2(X'X)^{-1}).$$

- So everything works out with inferences even if the error terms are not normally distributed as long as the sample size is reasonable large.

4.3 Non-Stationarity

- Definition: Integrated Processes: If x_t is a stationary process, then y_t, where $\Delta y_t = x_t$, is said to be integrated of order 1, or $I(1)$. The order that a process is integrated is equal to exactly how many times you have to difference the process to make it stationary. Note Δ signifies the first difference so $\Delta y_t = y_t - y_{t-1}$. Since x_t is stationary, it is $I(0)$.

- If z_t satisfies $\Delta z_t = y_t$ then z_t is $I(2)$. Thus,

$$\underbrace{\Delta\Delta z_t}_{\Delta^2 z_t}=\Delta y_t = x_t \equiv \text{ Stationary.}$$

- Example: Random Walk. $\Delta y_t = \epsilon_t \Longrightarrow y_t - y_{t-1} = \epsilon_t \Longrightarrow y_t = \phi y_{t-1} + \epsilon_t$ with $\phi = 1$. Thus $\Delta y_t = \epsilon_t$ is an $AR(1)$ process. Consider the graphs in the notes for the next part. **[G-4.1]** We can think of $\epsilon_t \sim iid(0,\sigma^2)$ as "white noise." White noise fluctuates above and below the axis and crosses it several times over the lifetime of the process. It usually doesn't drift very far from zero.

- In a Random walk, the expected amount of time until the first crossing is infinite (though usually this happens very quickly!) and in general, though the expected value of the process is zero, the process only crosses the axis very infrequently.

- Consider the properties of $\Delta y_t = \epsilon_t$. Suppose $y_0 = 0$ so we start out at the origin. Thus,

$$y_1 = \epsilon_1.$$

$$y_2 = y_1 + \epsilon_2 = \epsilon_1 + \epsilon_2.$$

$$y_3 = y_2 + \epsilon_3 = \epsilon_1 + \epsilon_2 + \epsilon_3.$$

$$\vdots$$

$$y_t = \sum_{s=1}^{t} \epsilon_s.$$

Thus,

$$E[y_t] = E[\sum_{s=1}^{t} \epsilon_s] = 0.$$

(As we expected). And the variance:

$$Var[y_t] = Var[\epsilon_1 + \epsilon_2 + \epsilon_3 + \cdots + \epsilon_t].$$

Since the variance of a sum is the sum of the variances (since ϵ_t is *iid*),

$$Var[y_t] = t\sigma^2.$$

So y_t is clearly NOT stationary because its variance depends on t. See graph. [**G-4.2**] The distribution of y_t is always centered at 0 for all t, but as t gets larger, the distribution gets more spread out.

- Memory. Consider a stationary $AR(1)$ process: $y_t = \phi y_{t-1} + \epsilon_t$ with $|\phi| < 1$. Then we have shown:

$$y_t = \sum_{s=1}^{t} \phi^s \epsilon_{t-s}.$$

So as $s \longrightarrow \infty$, $\phi^s \longrightarrow 0$. Thus as time progresses, a shock in period t has less and less impact in subsequent periods. "The memory of this process fades away."

- Now consider a non-stationary random walk process with:

$$y_t = \sum_{s=1}^{t} \epsilon_s.$$

A shock at period t has equal impact on ALL subsequent periods. Shocks last forever. "This process has long memory."

5 Week 5: 11 Feb - 15 Feb

5.1 Random Walk with Drift

- Consider the following model.

$$y_t = \alpha + y_{t-1} + \epsilon_t, \quad \epsilon_t \sim \ iid.$$

- The addition of the constant term to this simple AR(1) (non-stationary) process changes it from a "Random Walk" to a "Random Walk with Drift." Suppose $y_0 = 0$. Thus, substituting:

$$y_t = \alpha + y_{t-1} + \epsilon_t.$$
$$y_t = \alpha + (\alpha + y_{t-2} + \epsilon_{t-1}) + \epsilon_t = 2\alpha + y_{t-2} + (\epsilon_t + \epsilon_{t-1}).$$
$$y_t = 2\alpha + (\alpha + y_{t-3} + \epsilon_{t-2}) + (\epsilon_t + \epsilon_{t-1}) = 3\alpha + y_{t-3} + (\epsilon_t + \epsilon_{t-1} + \epsilon_{t-2}).$$
$$\vdots$$
$$y_t = \underbrace{\alpha t}_{Trend/Drift} + \underbrace{\sum_{i=0}^{t-1} \epsilon_{t-i}}_{Random\ Walk}.$$

- Note that $E[y_t] = \alpha t$ and $Var[y_t] = t\sigma^2$. Notice the mean changes over time and the variance of the process gets larger. See graph in notes. **[G-5.1]**

5.1.1 Mean Shift

- Suppose we had the following model.

$$\begin{cases} y_t = \alpha_0 + \beta y_{t-1} + \epsilon_t & t \leq t_0 \\ y_t = \alpha_1 + \beta y_{t-1} + \epsilon_t & t > t_0 \end{cases} \tag{8}$$

- Because we have a mean change (from $\alpha_0 t$ to $\alpha_1 t$), even if $|\beta| < 1$, the process is non-stationary because $\alpha_0 \neq \alpha_1$.

5.2 Test for Stationarity - Test for the Unit Root

- Suppose we would like to test the following model for stationarity:

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

- For this process to be stationary, α must be zero and $|\phi| < 1$. We will proceed in two parts. First just checking to see if ϕ is in absolute value less than 1, and if we find that it is NOT, ie we have a random walk, then we'll check to see if α is not zero, ie, we have a random walk with drift.

- Note that testing for stationarity under the null hypothesis of stationarity will not work. Any process with $|\phi| < 1$ is stationary, even one in which $\phi = 0.99999$. To get this sort of precision, (distinguish between 1 and 0.99999), we need many many observations. Thus, we will use Non-Stationarity as our null hypothesis.

- Consider the maintained model:

$$y_t = \phi y_{t-1} + \epsilon_t \quad \epsilon_t \sim \ iid(0, \sigma^2).$$

- Hypotheses:

$$H_0 : |\phi| = 1 \quad (Non - Stationary).$$
$$H_1 : |\phi| < 1 \quad (Stationary).$$

- So the procedure would be to run the OLS regression above and get an estimate for $\hat{\phi}$. However, because the distribution of $\hat{\phi}$ under the maintained model is not asymptotically normal, we cannot use the t or F tables to do the usual test.

- Thus, we use the following test statistic and compare it to the Dickey-Fuller tables (middle panel) which are calculated using a boot-strapping process and are accurate to within reasonable limits. Compute:

$$\hat{\tau}_\mu = \frac{\hat{\phi} - 1}{\left[\frac{s^2}{\sum(y_{t-1} - \bar{y}_{t-1})^2}\right]^{1/2}}.$$

  A typical critical value for this test is around -3.0.

- So we now have a test to determine the value of ϕ, we would now like to extend the model to see if there should be an α term added on to make this a random walk with drift. Suppose we have established the hypothesis $|\phi| = 1$ cannot be rejected. The next step is to investigate $\hat{\alpha}$. Consider the maintained model with the additive term:

$$y_t = \alpha + \beta t + \phi y_{t-1} + \epsilon_t \quad \epsilon_t \sim \ iid(0, \sigma^2).$$

- Hypothesis:

$$H_0 : |\phi| = 1 \ and \ \beta = 0 \quad (Random\ Walk\ With\ Drift).$$

  So testing $|\phi| = 1$ alone has yielded a conclusion of non-stationarity. Here we test to see if including the time trend yeilds the same results. I'm still not sure why we don't test for $\alpha = 0$, but it has something to do with the test on the time trend being more powerful.

- Estimate:

$$\hat{\tau}_\tau = \frac{\phi - 1}{SE}.$$

Use the Bottom panel of the Dicky Fuller tables.

5.2.1 Higher Order AR Processes

- Consider at first the simple stationary AR(1) process:

$$x_t = \phi x_{t-1} + \epsilon_t, \quad |\phi| < 1.$$

- And now consider another process, y_t:

$$\Delta y_t = y_t - y_{t-1} = x_t.$$

Therefore y_t is intergrated of order 1: $I(1)$. Substituting,

$$\Delta y_t = \phi \Delta y_{t-1} + \epsilon_t.$$

Noting that x_t has root, ϕ, what are the roots of the y_t process? Consider rewriting as follows:

$$\Delta y_t = \phi \Delta y_{t-1} + \epsilon_t.$$
$$y_t - y_{t-1} = \phi(y_{t-1} - y_{t-2}) + \epsilon_t.$$

$$y_t - y_{t-1} - \phi y_{t-1} = -\phi y_{t-2} + \epsilon_t.$$

$$y_t - (1+\phi) y_{t-1} = -\phi y_{t-2} + \epsilon_t.$$

$$y_t - (1+\phi) y_{t-1} + \phi y_{t-2} = \epsilon_t.$$

Noting that coefficients in front of each term are: 1, $-(1+\phi)$, and ϕ, the roots of y_t, will be the roots of the following quadratic:

$$z^2 - (1+\phi)z + \phi = 0.$$

Factorizing,

$$(z-1)(z-\phi) = 0.$$

Thus, the roots are $z = 1$ and $z = \phi$.

Since $z = 1$ is one of the roots, y_t is non-stationary. To get stationarity, we would need $|z_i| < 1$ for all i.

- Consider a general $AR(p)$ process:

$$x_t = \phi_1 x_{t-1} + \phi_2 x_{t-2} + \phi_3 x_{t-3} + \cdots + \phi_p x_{t-p} + \epsilon_t.$$

Assume that x_t is stationary, ie the roots of x_t, $(z_1, z_2, \ldots, z_p)$ are all less than 1 in absolute value.

- Consider a second process:

$$\Delta y_t = x_t.$$

Thus,

$$\Delta y_t = \phi_1 \Delta y_{t-1} + \phi_2 \Delta y_{t-2} + \phi_3 \Delta y_{t-3} + \cdots + \phi_p \Delta y_{t-p} + \epsilon_t.$$

$y_t = y_{t-1} + x_t$ has roots: $1, z_1, z_2, \ldots, z_p$. Thus because one of the roots of y_t is 1, y_t is non-stationary.

- Suppose we have an $AR(p)$ process. How do we test for a unit root? Consider an $AR(p)$:

$$y_t = \alpha + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \phi_3 y_{t-3} + \cdots + \phi_p y_{t-p} + \epsilon_t.$$

Notice that the roots of this equation solve:

$$z^p - \phi_1 z^{p-1} - \phi_2 z^{p-2} - \cdots - \phi_{p-1} z - \phi_p = 0.$$

If $z = 1$ is one of the roots, we could plug it into this equation and solve. Thus, letting $z = 1$:

$$1^p - \phi_1 1^{p-1} - \phi_2 1^{p-2} - \cdots - \phi_{p-1} 1 - \phi_p = 0.$$

$$1 - \phi_1 - \phi_2 - \cdots - \phi_{p-1} - \phi_p = 0.$$

$$1 - \sum_{i=1}^{p} \phi_i = 0.$$

$$\sum_{i=1}^{p} \phi_i = 1.$$

So if the sum of the ϕ's exactly equals 1, we know that we have a unit root.

5.3 Augmented Dicky Fuller Unit Root Test

- We will now show that the following $AR(p)$ process:

$$y_t = \alpha + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \phi_3 y_{t-3} + \cdots + \phi_p y_{t-p} + \epsilon_t.$$

  Can be rewritten:

$$\Delta y_t = \alpha + \phi_0^* y_{t-1} + \sum_{j=1}^{p-1} \phi_j^* \Delta y_{t-j} + \epsilon_t.$$

  With:

$$\phi_0^* = -(1 - \sum_{i=1}^{p} \phi_i).$$

  And,

$$\phi_j^* = -\sum_{k=j+1}^{p} \phi_k.$$

- Consider and $AR(p)$ process:

$$y_t = \alpha + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \cdots + \phi_p y_{t-p} + \epsilon_t.$$

  We can rewrite this (by rejigging) by subtracting y_{t-1} from several terms and then adding them back in, in a special way:

$$y_t - y_{t-1} = \alpha + \left[\phi_1 - 1 + \phi_2 + \phi_3 + \cdots + \phi_p\right] y_{t-1} + \phi_2(y_{t-2} - y_{t-1})$$
$$+\phi_3(y_{t-3} - y_{t-1}) + \cdots + \phi_p(y_{t-p} - y_{t-1}) + \epsilon_t.$$

  Notice that $y_{t-2} - y_{t-1} = -(y_{t-1} - y_{t-2}) = -\Delta y_{t-1}$. Also, $y_{t-3} - y_{t-1} = -(y_{t-1} - y_{t-3}) = -(y_{t-1} - y_{t-2} + y_{t-2} - y_{t-3}) = -(\Delta y_{t-1} + \Delta y_{t-2})$. And so on. Thus, substituting in,

$$y_t - y_{t-1} = \alpha + \left[\sum_{i=1}^{p} \phi_i - 1\right] y_{t-1} - \phi_2 \Delta y_{t-1} - \phi_3(\Delta y_{t-1} + \Delta y_{t-2})$$
$$-\phi_4(\Delta y_{t-1} + \Delta y_{t-2} + \Delta y_{t-3}) - \cdots - \phi_p(\Delta y_{t-1} + \Delta y_{t-2} + \cdots + \Delta y_{t-p+1}) + \epsilon_t.$$

  Combining terms and rearranging:

$$\Delta y_t = \alpha - \left[1 - \sum_{i=1}^{p} \phi_i\right] y_{t-1} - (\phi_2 + \phi_3 + \cdots + \phi_p)\Delta y_{t-1} - (\phi_3 + \phi_4 + \cdots + \phi_p)\Delta y_{t-2} - \cdots - \phi_p \Delta y_{t-p+1} + \epsilon_t.$$

Which is identical to the result we were looking for:

$$\Delta y_t = \alpha + \phi_0^* y_{t-1} + \sum_{j=1}^{p-1} \phi_j^* \Delta y_{t-j} + \epsilon_t,$$

with:

$$\phi_0^* = -(1 - \sum_{i=1}^{p} \phi_i),$$

and,

$$\phi_j^* = - \sum_{k=j+1}^{p} \phi_k.$$

- Notice that y_{t-1} is the only level term and can be dated at any point in time but by convention, is usually labeled y_{t-1}.
- Thus, y_t has a unit root iff and coefficient on y_{t-1}, $1 - \sum_i \phi_i$, is equal to zero. Or, via last lecture's result that y_t has a unit root iff $\sum_i \phi_i = 1$. The procedure is thus to regress Δy_t on a constant, y_{t-1}, Δy_{t-1}, $\Delta y_{t-2}, \ldots, \Delta y_{t-p+1}$. Then calculate:

$$\hat{\tau}_\mu = \frac{\widehat{\text{Coefficient on } y_{t-1}}}{SE},$$

  and compare it to the Dicky Fuller tables, middle panel. This is called the Augmented Dicky Fuller test or ADF test.
- For yearly data, it is probably sufficient to only include y_{t-1} and Δy_{t-1}. For quarterly data, it is usually best to use about 4 or 5 lags, however the best approach is to "Suck it and see."

5.4 Detrending

- Linear Detrending: Take y_t and regress it on a constant and a time trend. Take the residuals. Note, if we detrend y_t and x_t to obtain the detrended processes: y_t^* and x_t^*, and then we regress y_t^* on x_t^* and a constant, this gives the same answer as regressing y_t on x_t, a constant AND a time trend.
- Hodrick - Prescott Filter (HP Filter). If y_t is a series and you want to find a detrended series, y_t^*, then y_t^* is chose to solve the following equation:

$$Min_{y_1^*,\ldots,y_T^*} \quad \underbrace{\sum_{t=1}^{T} (y_t^* - y_t)^2}_{\text{Closeness to } y_t} + \lambda \underbrace{\sum_{t=2}^{T} (y_t^* - y_{t-1}^*)^2}_{\text{Smoothness}}.$$

Notice that λ is the control term that determines how much weight you want to put on the smoothness of your new detrended series. If $\lambda = 0$, then $y_t^* = y_t$ and you haven't done anything. Higher values of λ correspond to smoother fits. $\lambda = 1600$ is conventional.

- See notes for graphs that compare linear detrending to HP detrending. **[G-5.2]** The HP filter is slightly more sophisticated and thus can fit a wider range of data patterns.

5.5 Models where Errors follow an AR process

- Consider the following model:

$$y_t = x_t'\beta + u_t, \quad t = 1 \dots T.$$

However, u_t is NOT iid. Instead,

$$u_t = \phi u_{t-1} + \epsilon_t, \quad \epsilon_t \sim \; iid(0, \sigma^2), \quad |\phi| < 1.$$

- Recall:

$$Var(u_t) = \frac{\sigma^2}{1-\phi^2}.$$

$$Cov(u_t, u_{t-s}) = \frac{\phi^s \sigma^2}{1-\phi^2}.$$

Thus,

$$V(u) = E[uu'] = \frac{\sigma^2}{1-\phi^2} \begin{bmatrix} 1 & \phi & \phi^2 & \dots & \phi^{T-1} \\ \phi & \ddots & \ddots & \ddots & \vdots \\ \phi^2 & \ddots & \ddots & \ddots & \phi^2 \\ \vdots & \ddots & \ddots & \ddots & \phi \\ \phi^{T-1} & \dots & \phi^2 & \phi & 1 \end{bmatrix}. \tag{9}$$

Call this $E[uu'] = \sigma^2 \Omega$ where $\Omega = \frac{1}{1-\phi^2}[\,\cdot\,]$.

- Now suppose that ϵ_t is process independent of x_t. ie, x_t is uncorrelated with and independent of all ϵ. This implies x_t is independent of u_t because u_t is a function of all past ϵ such that:

$$u_t = \sum_{i=1}^{T} \phi^i \epsilon_{t-i}.$$

This means that OLS will generate consistant and unbiased estimates of β. However, the usual estimates of $Var(\hat{\beta})$ $(= \sigma^2(X'X)^{-1})$, are wrong. For example if X is non-stochastic,

$$Var(\hat{\beta}) = E[(\hat{\beta}-\beta)(\hat{\beta}-\beta)'] = E[(X'X)^{-1}X'u{\cdot}u'X(X'X)^{-1}] = (X'X)^{-1}X'\Omega X(X'X)^{-1}.$$

But this is as far as we can go. Usually the estimate, $\sigma^2(X'X)^{-1}$ will underestimate the true variance just derived. Thus coefficients will look more significant then they really are.

6 Week 6: 18 Feb - 22 Feb

6.1 Models with Serially Correlated Error Terms

- Consider the following model:

$$y_t = \gamma y_{t-1} + x_t'\beta + u_t, \quad |\gamma| < 1.$$

Assume that the x_t's are stationary and independent of u_s, for all t and s. In terms of error, then:

$$u_t = \phi u_{t-1} + \epsilon_t, \quad \epsilon_t \sim \; iid(0, \sigma^2), \quad |\phi| < 1.$$

- Because u_t is stationary and we have assumed that x_t is stationary, then y_t is also stationary.

- This model will fail under ordinary OLS because of the nature of the error terms. Namely, the serial correlation of u_t, makes $Cov(y_{t-1}, u_t) \neq 0$. Investigate this by considering the expectation between the two terms:

$$E[y_{t-1}u_t] = E[(\gamma y_{t-2} + x_{t-1}'\beta + u_{t-1})(\phi u_{t-1} + \epsilon_t)].$$

Mulitplying out,

$$= \gamma\phi E[y_{t-2}u_{t-1}] + \gamma E[y_{t-2}\epsilon_t] + \phi E[x_{t-1}'u_{t-1}]\beta + E[x_{t-1}'\epsilon_t]\beta + \phi E[u_{t-1}^2] + E[u_{t-1}\epsilon_t].$$

Noting that all terms in expectation with ϵ_t that are dated before time t are zero because ϵ_t is iid,

$$E[y_{t-1}u_t] = \gamma\phi E[y_{t-2}u_{t-1}] + \phi E[u_{t-1}^2].$$

Since y_t is stationary, $E[y_{t-1}u_t] = E[y_{t-2}u_{t-1}]$. Thus,

$$E[y_{t-1}u_t] - \gamma\phi E[y_{t-2}u_{t-1}] = \phi E[u_{t-1}^2].$$
$$E[y_{t-1}u_t] - \gamma\phi E[y_{t-1}u_t] = \phi E[u_{t-1}^2].$$
$$E[y_{t-1}u_t](1 - \gamma\phi) = \phi E[u_{t-1}^2].$$

Now, we know that for a standard $AR(1)$ process, like u_t, $E[u_{t-1}^2] = \dfrac{\sigma^2}{1-\phi^2}$. Thus,

$$E[y_{t-1}u_t](1 - \gamma\phi) = \phi\frac{\sigma^2}{1-\phi^2}.$$

Or finally,

$$E[y_{t-1}u_t] = \frac{\sigma^2\phi}{(1-\phi^2)(1-\gamma\phi)} \neq 0.$$

- So y_{t-1} and u_t are correlated and OLS yields inconsistent estimators.

6.1.1 Estimation with Serially Correlated Error

- Consider the following model (as above):

$$y_t = x_t'\beta + u_t, \qquad u_t = \phi u_{t-1} + \epsilon_t, \quad |\phi| < 1.$$

- We have shown that the variance covariance matrix of the errors is as follows:

$$V(u) = E[uu'] = \frac{\sigma^2}{1-\phi^2}\begin{bmatrix} 1 & \phi & \phi^2 & \dots & \phi^{T-1} \\ \phi & \ddots & \ddots & \ddots & \vdots \\ \phi^2 & \ddots & \ddots & \ddots & \phi^2 \\ \vdots & \ddots & \ddots & \ddots & \phi \\ \phi^{T-1} & \dots & \phi^2 & \phi & 1 \end{bmatrix}. \tag{10}$$

Call this $E[uu'] = \sigma^2\Omega$ where $\Omega = \frac{1}{1-\phi^2}[\,\cdot\,]$.

- If the errors are heterscedastic (meaning that $\phi \neq 0$), we must use generalized least squares instead of OLS. Suppose ϕ is known. We could just estimate as follows:

$$\hat{\beta}_{GLS} = (X'\Omega^{-1}X)^{-1}X'\Omega^{-1}y.$$
$$Var(\hat{\beta}_{GLS}) = \sigma^2(X'\Omega^{-1}X)^{-1}.$$

- Take the following two points on faith though they are fairly easy to verify.

$$\Omega^{-1} = \begin{bmatrix} 1 & -\phi & 0 & \dots & 0 \\ -\phi & (1+\phi^2) & \ddots & \ddots & \vdots \\ 0 & \ddots & \ddots & \ddots & 0 \\ \vdots & \ddots & \ddots & (1+\phi^2) & -\phi \\ 0 & \dots & 0 & -\phi & 1 \end{bmatrix}. \tag{11}$$

And Ω^{-1} can be written as $L'L$ where L is defined as:

$$L = \begin{bmatrix} (1-\phi^2)^{1/2} & 0 & \dots & \dots & 0 \\ -\phi & 1 & \ddots & \ddots & \vdots \\ 0 & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & 0 \\ 0 & \dots & 0 & -\phi & 1 \end{bmatrix}. \tag{12}$$

- We can use the fact that $\Omega^{-1} = L'L$ in the following situation. Suppose we look at the model:

$$Ly = LX\beta + Lu.$$

And suppose $y^* = Ly$ and $X^* = LX$. Note first the terms of Ly:

$$\begin{aligned} y_1^* &= (1-\phi^2)^{1/2} y_1. \\ y_2^* &= y_2 - \phi y_1. \\ &\vdots \\ y_t^* &= y_t - \phi y_{t-1}. \end{aligned}$$

And the elements of LX:

$$\begin{aligned} x_1^* &= (1-\phi^2)^{1/2} x_1. \\ x_2^* &= x_2 - \phi x_1. \\ &\vdots \\ x_t^* &= x_t - \phi x_{t-1}. \end{aligned}$$

- Now suppose we run OLS. Suppose we regress Ly on LX. The definition of $\hat{\beta}$ is as follows:

$$\hat{\beta} = ((LX)'(LX))^{-1}(LX)'Ly = (X'L'LX)^{-1}X'L'Ly = (X'\Omega^{-1}X)^{-1}X'\Omega^{-1}y = \hat{\beta}_{GLS}.$$

So running the regression of Ly and LX is equivalent to running the GLS regression.

- The Ly and LX transformation above is called the Cochran-Orcutt (C-O) transformation.

- Consider the errors in our regression, $Ly = LX\beta + Lu$.

$$\begin{aligned} Var(Lu) &= E[Lu(Lu)'] = E[Luu'L'] = LE[uu']L' = L\sigma^2\Omega L' \\ &= \sigma^2 L(L'L)^{-1}L' = \sigma^2 LL^{-1}L'^{-1}L' = \sigma^2 I. \end{aligned}$$

So the variances are now homoscedastic once we transform with C-O.

- To see this another way, consider the model again,

$$y_t = x_t'\beta + u_t, \qquad u_t = \phi u_{t-1} + \epsilon_t.$$

Now take the y_t process, lag it once and muliply by ϕ,

$$\phi y_{t-1} = \phi x_{t-1}'\beta + \phi u_{t-1}.$$

Subtract $y_t - \phi y_{t-1}$:

$$y_t - \phi y_{t-1} = (x_t' - \phi x_{t-1}')\beta + u_t - \phi u_{t-1}.$$

$$y_t - \phi y_{t-1} = (x_t' - \phi x_{t-1}')\beta + \underbrace{\epsilon_t}_{IID!!}.$$

Which makes the regression yield consistant estimators.

- However usually ϕ is NOT known. The obvious solution would be to estimate ϕ and use the estimate as above in the C-O transformation. There are several methods for doing so:

- Method 1.

  - Works for process independent x's. First do OLS on $y = X\beta + u$. Note that β will be consistent but the standard errors will all be incorrect. Then save the residuals from this regression and run another regression of u_t on u_{t-1} to find an estimate of ϕ such that:

$$\hat{\phi} = \frac{\sum_{t=2}^{T} \hat{u}_t \hat{u}_{t-1}}{\sum_{t=2}^{T} \hat{u}_{t-1}^2}.$$

  - Then do the C-O transformation and use OLS to find the solution.
  - Note: this does not work for x only contemporaneously independent. Why? Because the first step will yield inconsistent estimates of β which yields inconsistant estimates of u and therefore ϕ.

- Method 2. (Durbin)

  - Write model in transformed form, i.e.,

$$y_t - \phi y_{t-1} = (x_t' - \phi x_{t-1}')\beta + \epsilon_t.$$

$$y_t = \phi y_{t-1} + x_t'\beta - \phi x_{t-1}'\beta + \epsilon_t.$$

  - Estimate this equation by OLS and the coefficient on y_{t-1}, $\hat{\phi}$, will be a consistent estimate of ϕ. Use $\hat{\phi}$ to do a C-O transformation procedure and then run OLS to find an estimate of β.

6.2 More Methods of Estimation with Serially Correlated Errors

- Method 2.
  - Consider the following model:

$$y_t = \gamma y_{t-1} + x_t'\beta + u_t, \quad |\gamma| < 1.$$

    With errors:

$$u_t = \phi u_{t-1} + \epsilon_t, \quad |\phi| < 1, \quad \epsilon_t \sim \ iid.$$

  - Perform the transformation as we did in last section by lagging y_t once, multiplying everything through by ϕ and subtracting the result from y_t. Thus,

$$y_t - \phi y_{t-1} = \gamma y_{t-1} - \phi\gamma y_{t-2} + x_t'\beta - x_{t-1}'\beta\phi + \epsilon_t.$$

    Or rewritting,

$$y_t = (\phi + \gamma)y_{t-1} - \phi\gamma y_{t-2} + x_t'\beta - x_{t-1}'\beta\phi + \epsilon_t.$$

    Note that $\epsilon_t = u_t - \phi u_{t-1}$.
  - Our aim is still to determine the coefficient on y_{t-1}, γ. However, if we look at the coefficients on the two terms, y_{t-1} and y_{t-2}, we have two equations and two unknowns, but the problem will be that we don't know which is γ and which is ϕ. Note also that we can't utilize the coefficients on x_t' and x_{t-1}' because β is a vector and we can't just divide the coefficients to find the value of γ.
  - Thus, method 3.

- Method 3. Non-Linear Least Squares.
  - This is basically maximum likelihood estimation.
  - Suppose first of all that $\epsilon_t \sim N(0, \sigma^2)$. From the above equation defining y_t, we find that the density of y_t conditional on y_{t-1}, y_{t-2}, x_t', and x_{t-1}' is distributed:

$$N((\phi + \gamma)y_{t-1} - \phi\gamma y_{t-2} + x_t'\beta - x_{t-1}'\beta\phi, \sigma^2).$$

  - This leads to the following log-likelihood function:

$$Log\ L(\phi, \gamma, \beta, \sigma^2) = -\frac{-(T-2)}{2}log(2\pi) - \frac{-(T-2)}{2}log(\sigma^2) - \frac{1}{2\sigma^2}\sum_{t=3}^{T}\epsilon_t^2.$$

    Where $\epsilon_t = y_t - (\phi + \gamma)y_{t-1} + \phi\gamma y_{t-2} - x_t'\beta + x_{t-1}'\beta\phi$. Note that the summation goes from $t = 3$ to T because we need the first two lags to calculate the first y_t.

– To solve this maximization, we will need to consider the first order condition of *Log L*. First of all calculate z_t as follows:

$$z_t = \begin{bmatrix} -\frac{\partial \epsilon_t}{\partial \beta} \\ -\frac{\partial \epsilon_t}{\partial \phi} \\ -\frac{\partial \epsilon_t}{\partial \gamma} \end{bmatrix} = \begin{bmatrix} x_t' - \phi x_{t-1}' \\ y_{t-1} - \gamma y_{t-2} - x_{t-1}'\beta \\ y_{t-1} - \phi y_{t-2} \end{bmatrix}. \quad (13)$$

Thus the FOC will be $\sum_{t=3}^{T} z_t \epsilon_t = 0$.

6.3 Hypothesis Testing in the ML context

- 3 tests: Wald, Likelihood Ratio, and the Lagrange multiplier test.

6.3.1 The Wald Test

- Suppose we have ψ parameters and log likelihood, $log\ L(\psi)$. The ML estimate of ψ, namely, $\hat{\psi}$, has the property,

$$\sqrt{T}(\hat{\psi} - \psi) \longrightarrow^{D} N(0, IA(\psi)^{-1}).$$

Where IA is the asymptotic information matrix such that $IA(\psi) = lim\frac{1}{T}I(\psi)$ where $I(\psi)$ is the information matrix, such that,

$$I(\psi) = -E\left[\frac{\partial^2 log\ L(\psi)}{\partial\psi\partial\psi'}\right].$$

- We construct the hypotheses as follows:

$$H_0 : R\psi - r = 0.$$
$$H_1 : R\psi - r \neq 0.$$

Where R is a matrix of q ($< k$) restrictions (ie, R has q linearly independent rows and k is the dimension of ψ).

- Consider an example. Suppose we want to test the following two restrictions when ψ is $1x3$:

$$H_0 : \psi_1 + \psi_2 = 1, \psi_1 - \psi_3 = 3.$$

We can write this in matrix form as follows:

$$R\psi - r = 0 \Longleftrightarrow \begin{bmatrix} 1 & 1 & 0 \\ 1 & 0 & -1 \end{bmatrix} \begin{bmatrix} \psi_1 \\ \psi_2 \\ \psi_3 \end{bmatrix} - \begin{bmatrix} 1 \\ 3 \end{bmatrix} = 0. \tag{14}$$

- Under H_0,

$$\sqrt{T}(R\hat{\psi} - r) = \sqrt{T}R(\hat{\psi} - \psi).$$

From above we therefore know,

$$\sqrt{T}R(\hat{\psi} - \psi) \longrightarrow^{D} N(0, R\, IA(\psi)^{-1}\, R').$$

- Thus noting that if $x \sim N(0, A)$, then $x'A^{-1}x \sim \chi^2$, we can write the Wald statistic as follows:

$$W = \sqrt{T}(R\hat{\psi} - r)' \Bigg[R\, IA(\psi)^{-1}\, R' \Bigg]^{-1} \sqrt{T}(R\hat{\psi} - r) \longrightarrow^{D} \chi^2(q).$$

- To use this test, replace $IA(\psi)$ with one of the following:
  - 1) $\frac{1}{T}I(\hat{\psi})$.
  - 2) $-\frac{1}{T}\Big[\dfrac{\partial log\ L(\psi)}{\partial\psi'\partial\psi}\Big]$.

- So if we replace it with the first case,

$$W = (R\hat{\psi} - r)' \Bigg[R\, I(\hat{\psi})^{-1}\, R') \Bigg]^{-1} (R\hat{\psi} - r) \longrightarrow^{D} \chi^2(q).$$

- To do the test, compute W and if W is significantly greater than 0 (recall that chi-squared is a positive distribution), then reject the null and conclude the restrictions embodied in R are NOT valid.

- The Wald test is simple to use and only requires estimating the unrestricted model. However it is very difficult to test non-linear restrictions using the Wald Test. Thus we move to the second test.

6.3.2 Likelihood Ratio Test

- Suppose we wanted to test non-linear restrictions such as $\psi_1\psi_2 = 7$ and $\psi_1 - \psi_3^2 = 4$. Write these two restrictions in a matrix R. To do this test, we couldn't use a Wald test, or at least it would be fairly difficult to do so because we would have to linearize the restrictions.

- So run a Likelihood Ratio (LR) test as follows:

- Compute $\hat{\psi}$ which solves

$$Max_{\psi} Log\ L(\psi).$$

  This is the UNRESTRICTED estimator.

- Compute $\hat{\psi}_0$ which solves

$$Max_{\psi} Log\ L(\psi)$$

  Subject to:

$$R(\psi) = 0.$$

  This is the RESTRICTED estimator.

- In general $log\ L(\hat{\psi}) > log\ L(\hat{\psi}_0)$. Compute the LR:

$$LR = -2log\left[\frac{L(\hat{\psi}_0)}{L(\hat{\psi})}\right] = 2\left[logL(\hat{\psi}) - logL(\hat{\psi}_0)\right].$$

  LR is aymptotically distributed $\chi^2(q)$ where q is the number of restrictions in R.

7 Week 7: 25 Feb - 1 Mar

7.1 The Trinity of Tests - Continued

7.1.1 The Lagrange Multiplier Test - LM Test

- In this test, we only have to estimate the restricted model which is usually easier to do. Note that in the Wald test, we only had to estimate the unrestricted and in the Likelihood ratio test, we needed to estimate both.

- The test is as follows. Suppose we have the likelihood function formed as usual, $L(\psi)$. We also have a set of restrictions of the form:

$$R(\psi) = 0.$$

  Let $\hat{\psi}_0$ be the ML estimate of the restricted model. The FOC of the unrestricted model is:

$$\frac{\partial Log\ L(\psi)}{\partial \psi} = 0.$$

- If the restrictions are close to being satisfied, then

$$\frac{\partial Log\ L(\hat{\psi}_0)}{\partial \psi}$$

  should be close to zero.

- Then the LM test is based on:

$$\frac{1}{T} \frac{\partial Log\ L(\hat{\psi}_0)}{\partial \psi'} IA(\psi)^{-1} \frac{\partial Log\ L(\hat{\psi}_0)}{\partial \psi} \longrightarrow^{D} \chi^2(q).$$

  Where q is the number of restrictions.

- Replacing the Asymptotic Information matrix, $IA(\psi)$, by an estimate, eg: $\frac{1}{T}I(\hat{\psi})$, gives the LM statistic:

$$LM = \frac{\partial Log\ L(\hat{\psi}_0)}{\partial \psi'} I(\hat{\psi})^{-1} \frac{\partial Log\ L(\hat{\psi}_0)}{\partial \psi} \longrightarrow^{a} \chi^2(q).$$

  Where the "a" above the arrow means "Converges Asymptotically in Distribution."

- Consider using the LM test in Non-linear least squares models. Note that to compute the LM statistic, we will need the partial derivative of the log likelihood function evaluted at the restricted coefficents and also the information matrix. Consider the general $N.L.L.S.$ model:

$$y_t = g(x_t; \beta) + \epsilon_t, \quad \epsilon_t \sim iid\ N(0, \sigma^2).$$

Also assume that x_t and ϵ_t are independent.

- The log likelihood function is therefore:

$$Log\ L(\beta, \sigma^2) = -\frac{T}{2} Log(2\pi) - \frac{T}{2} Log(\sigma^2) - \frac{1}{2\sigma^2} \sum_{t=1}^{T} \epsilon_t(\beta)^2.$$

With,

$$\epsilon_t(\beta) = y_t - g(x_t; \beta).$$

- Suppose we have a set of restrictions on β of the form, $R\beta = 0$. The FOC is therefore:

$$\frac{\partial Log\ L}{\partial \beta} = -\frac{1}{\sigma^2} \sum_{t=1}^{T} \epsilon_t \frac{\partial \epsilon_t}{\partial \beta}.$$

Let $z_t = -\frac{\partial \epsilon_t}{\partial \beta} = \frac{\partial g(x_t; \beta)}{\partial \beta}$. Thus,

$$\frac{\partial Log\ L}{\partial \beta} = \frac{1}{\sigma^2} \sum_{t=1}^{T} \epsilon_t z_t.$$

- Now for the information matrix. In this case the I matrix is block diagonal and the block corresponding to β is:

$$\frac{1}{\sigma^2} \sum_{t=1}^{T} E[z_t z_t'].$$

- So if the restricted estimates are $\hat{\beta}_0$ and $\hat{\sigma}_0^2$, then our statistic is as follows (plugging in the partial and the information matrix):

$$LM = \left(\frac{1}{\hat{\sigma}_0^2} \sum_{t=1}^{T} \epsilon_t(\hat{\beta}_0) z_t(\hat{\beta}_0) \right)' \left(\frac{1}{\hat{\sigma}_0^2} \sum_{t=1}^{T} E[z_t(\hat{\beta}_0) z_t(\hat{\beta}_0)'] \right)^{-1} \left(\frac{1}{\hat{\sigma}_0^2} \sum_{t=1}^{T} \epsilon_t(\hat{\beta}_0) z_t(\hat{\beta}_0) \right).$$

Asymptotically, we can replace $\sum_{t=1}^{T} E[z_t z_t']$ with $\sum_{t=1}^{T} z_t z_t'$. Thus the simplified result:

$$LM = \frac{1}{\hat{\sigma}_0^2} \left(\sum_{t=1}^{T} \epsilon_t z_t \right)' \left(\sum_{t=1}^{T} z_t z_t' \right)^{-1} \left(\sum_{t=1}^{T} \epsilon_t z_t \right).$$

- This rather complicated expression can be found another way however. Suppose we find ϵ_t and calculate z_t and then regress ϵ_t on z_t. In other words we run the regression:

$$\epsilon_t = z_t'\gamma + \nu_t.$$

Thus, the standard OLS result:

$$\hat{\gamma} = (\sum z_t z_t')^{-1} \sum z_t \epsilon_t.$$

Note the fitted equation is therefore:

$$\hat{\epsilon}_t = z_t'\hat{\gamma}.$$

Substituting in for $\hat{\gamma}$,

$$\hat{\epsilon}_t = z_t'(\sum z_t z_t')^{-1} \sum z_t \epsilon_t.$$

And since the motivation here is to find the residual sum of squares, we would like to square this last expression to find RSS. Thus,

$$\begin{aligned}
\hat{\epsilon}_t'\hat{\epsilon}_t &= \left(z_t'(\sum z_t z_t')^{-1} \sum z_t \epsilon_t \right)' z_t'(\sum z_t z_t')^{-1} \sum z_t \epsilon_t. \\
&= (\sum z_t \epsilon_t)'(\sum z_t z_t')^{-1} z_t z_t' (\sum z_t z_t')^{-1} \sum z_t \epsilon_t.
\end{aligned}$$

Summing:

$$\sum(\hat{\epsilon}_t'\hat{\epsilon}_t) = (\sum z_t \epsilon_t)' \underbrace{(\sum z_t z_t')^{-1} \sum z_t z_t'}_{I} (\sum z_t z_t')^{-1} (\sum z_t \epsilon_t).$$

Simplifying:

$$\sum(\hat{\epsilon}_t'\hat{\epsilon}_t) = (\sum z_t \epsilon_t)'(\sum z_t z_t')^{-1}(\sum z_t \epsilon_t).$$

Amazingly, this is the numerator of the LM statistic. Next, note that:

$$\frac{1}{T}\sum \epsilon_t^2 = \hat{\sigma}_0^2.$$

The R^2 value from our regression of ϵ_t on z_t is $\frac{\sum \hat{\epsilon}_t^2}{\sum \epsilon_t^2}$. Thus,

$$T \cdot R^2 = \frac{\sum \hat{\epsilon}_t^2}{1/T \sum \epsilon_t^2} = \frac{(\sum z_t \epsilon_t)'(\sum z_t z_t')^{-1}(\sum z_t \epsilon_t)}{\hat{\sigma}_0^2} = LM.$$

So just by finding ϵ_t and z_t and regressing, we can easily find the LM statistic.

7.1.2 LM Test - Example

- In this example, we will use the LM test to test for serial correlation in the error terms. Consider the model:

$$y_t = x_t'\beta + u_t, \quad u_t = \phi u_{t-1} + \epsilon_t, \quad |\phi| < 1, \quad \epsilon_t \sim \ iid\ N(0, \sigma^2).$$

  [Note that although we assume normality, this test would still work asymptotically no matter what the distribution.]

- Test:

$$H_0 : \phi = 0 \ \text{(Errors Not Serially Correlated)}.$$
$$H_0 : \phi \neq 0 \ \text{(Errors Serially Correlated)}.$$

- Here, the LM test is good because we only need to estimate the restricted model which assumes the model does not suffer from serially correlated errors. Therefore regular OLS will provide consistent estimators.

- Follow these steps to compute LM:

- Step 1:

  - First calculate ϵ_t and z_t. Lagging y_t once and multiplying through by ϕ,

$$\phi y_{t-1} = \phi x_{t-1}'\beta + \phi u_{t-1}.$$

    Subtracting ϕy_{t-1} from y_t,

$$y_t - \phi y_{t-1} = x_t'\beta - \phi x_{t-1}'\beta + \underbrace{u_t - \phi u_{t-1}}_{\epsilon_t}.$$

$$y_t - \phi y_{t-1} = x_t'\beta - \phi x_{t-1}'\beta + \epsilon_t.$$

    Solving for ϵ_t,

$$\epsilon_t = y_t - \phi y_{t-1} - x_t'\beta + x_{t-1}'\beta\phi.$$

  - Recall that $z_t = -\dfrac{\partial \epsilon_t}{\partial \psi}$. Thus,

$$z_t = -\frac{\partial \epsilon_t}{\partial \psi} = \begin{bmatrix} -\dfrac{\partial \epsilon_t}{\partial \beta} \\ -\dfrac{\partial \epsilon_t}{\partial \phi} \end{bmatrix} = \begin{bmatrix} x_t' - \phi x_{t-1}' \\ y_{t-1} - x_{t-1}'\beta \end{bmatrix}. \tag{15}$$

  - Now calculate:

$$\frac{\partial log\ L}{\partial \psi} = \sum \epsilon_t z_t.$$

    in the unrestricted model.

- Step 2:
  - Compute the ML estimates of the parameters if the restriction holds, ie, if $\phi = 0$. Thus we get the OLS estimate for β, call it $\hat{\beta}_0$, and from the hypothesis, we have $\hat{\phi}_0 = 0$.
- Step 3:
  - Evaluate ϵ_t and z_t at these two parameters, $\hat{\beta}_0$ and $\hat{\phi}_0 = 0$. Thus,

$$\epsilon_t = y_t - x_t'\hat{\beta}_0.$$

    And,

$$z_t = \begin{bmatrix} x_t' \\ y_{t-1} - x_{t-1}'\hat{\beta}_0 \end{bmatrix}. \quad (16)$$

- Step 4:
  - Run regression of

$$\epsilon_t = \underbrace{x_t' \text{ and } y_{t-1} - x_{t-1}'\hat{\beta}_0}_{z_t}.$$

  - Take R^2 from this regression and compute:

$$LM = T \cdot R^2 \sim^a \chi^2(1).$$

  - Compare this statistic to the chi-squared distribution with 1 degree of freedom. Large values of LM will lead to a rejection of the null and a conclusion of serial correlation.

7.2 Durbin-Watson Test

- The DW test is for testing serial correlation among the residuals. Consider the following model:

$$y_t = x_t'\beta + u_t, \quad u_t = \phi u_{t-1} + \epsilon_t, \quad \epsilon_t \sim \ iid(0, \sigma^2).$$

  Test:

$$H_0 : \phi = 0.$$

- The DW test is only valid if all the x variables are process independent. It is not valid for example, if there are lagged dependent variables on the right hand side of the equation. Thus, is most instances, the DW test is not applicable.

- The Test Statistic:

$$d = \frac{\sum_{t=2}^{T}(\hat{u}_t - \hat{u}_{t-1})^2}{\sum_{t=1}^{T}\hat{u}_t^2}.$$

Multiplying out the top and seperating the terms,

$$d = \underbrace{\frac{\sum_{t=2}^{T}\hat{u}_t^2}{\sum_{t=1}^{T}\hat{u}_t^2}}_{\approx 1} - 2\underbrace{\frac{\sum_{t=2}^{T}\hat{u}_t\hat{u}_{t-1}}{\sum_{t=1}^{T}\hat{u}_t^2}}_{\approx\hat{\phi}} + \underbrace{\frac{\sum_{t=2}^{T}\hat{u}_{t-1}^2}{\sum_{t=1}^{T}\hat{u}_t^2}}_{\approx 1}.$$

Thus,

$$d \approx 1 - 2\hat{\phi} + 1 = 2 - 2\hat{\phi} = 2(1-\hat{\phi}).$$

- So when $\phi = 0$, ie we do not have serially correlated errors, the value of the DW statistic will be approximately 2. Note the range of d is $(0, 4)$ as $\hat{\phi}$ goes from $+1$ to -1. The DW tables list both an upper bound for d, d_U, and a lower bound for d, d_L. The decision rule is:

$$\text{if } d < d_L, \text{ then reject.}$$
$$\text{if } d > d_H, \text{ then do not reject.}$$
$$\text{if } d_L \le d \le d_H, \text{ then ambiguous.}$$

7.3 Modeling Economic Relationships

- The basic empirical model is a stochastic difference equation:

$$A(L)y_t = B(L)x_t + \epsilon_t, \quad \epsilon_t \sim \ iid(0, \sigma^2).$$

Where,

$$A(L) = 1 - a_1L - a_2L^2 - a_3L^3 - \cdots - a_nL^n.$$

And,

$$B(L) = b_0 + b_1L + b_2L^2 + b_3L^3 + \cdots + b_nL^n.$$

- Consider the following underlying models.

7.3.1 Adaptive Expectations Model

- Consider the model:

$$y_t = \beta x^e_{t+1} + \epsilon_t, \quad \epsilon_t \sim \ iid(0, \sigma^2).$$

- x^e_{t+1} is obviously not observed, so in this setting we say expectations are formed adaptively such that:

$$x^e_{t+1} = x^e_t + \gamma \underbrace{(x_t - x^e_t)}_{Mistake}, \quad \gamma \in [0, 1].$$

So here expectations of the level of x at time $t+1$ are based on what our expections were regarding today's level of x adjusted by how much we miss-judged the level.

- If $\gamma = 0$, then there is no adaption at all and $x^e_{t+1} = x^e_t$.
- If $\gamma = 1$, then we have instant adaptation and $x^e_{t+1} = x_t$.

  But is this type of process "rational"? This leads us the next model.

7.3.2 Rational Expectations Model

- Rational expectations means that $x^e_{t+1} = E[x_{t+1}|I_t]$. Here the information matrix at time t includes both ϵ_t and x_t.
- We will now show that the adaptive process above:

$$x^e_{t+1} = x^e_t + \gamma(x_t - x^e_t), \quad \gamma \in [0, 1]$$

is "Rational" iff x follows the process:

$$x_{t+1} = x_t + \epsilon_{t+1} - (1 - \gamma)\epsilon_t.$$

Note this is the equation of the x_t process itself, not the expectations equation. Lagging the process once,

$$x_t = x_{t-1} + \epsilon_t - (1 - \gamma)\epsilon_{t-1}.$$

Rewritting,

$$x_t - \epsilon_t = x_{t-1} - (1 - \gamma)\epsilon_{t-1}.$$

- Taking expectations of the x process noting that our best guess at ϵ_{t+1} is its expected value, zero:

$$x^e_{t+1} = x_t - (1 - \gamma)\epsilon_t.$$

Lagging this equation once:

$$x_t^e = x_{t-1} - (1-\gamma)\epsilon_{t-1}.$$

- Substituting in above from the lagged rewritten process:

$$x_t^e = x_t - \epsilon_t.$$

Multiplying through by $(1-\gamma)$,

$$(1-\gamma)x_t^e = (1-\gamma)x_t - (1-\gamma)\epsilon_t.$$

- Subtracting this from the expectation of the process,

$$x_{t+1}^e - (1-\gamma)x_t^e = x_t - (1-\gamma)x_t \underbrace{-(1-\gamma)\epsilon_t - (-(1-\gamma)\epsilon_t)}_{0} = x_t - (1-\gamma)x_t.$$

- Finally rewriting,

$$\begin{aligned} x_{t+1}^e - (1-\gamma)x_t^e &= x_t - (1-\gamma)x_t. \\ x_{t+1}^e &= (1-\gamma)x_t^e + x_t - (1-\gamma)x_t. \\ x_{t+1}^e &= x_t^e - \gamma x_t^e + \gamma x_t. \\ x_{t+1}^e &= x_t^e + \gamma(x_t - x_t^e). \end{aligned}$$

Which is exactly the model we had in the adaptive expectations model. So if x_t evolves as stated above, then the adaptive expectations model is "rational."

- Note that we can also write the adaptive expectations model using the lag operator as follows:

$$x_{t+1}^e = (1-\gamma)Lx_{t+1}^e + \gamma x_t.$$

Solving for x_{t+1}^e,

$$x_{t+1}^e = \frac{\gamma x_t}{1-(1-\gamma)L}.$$

Thus, plugging in, the final model becomes:

$$y_t = \beta\left[\frac{\gamma x_t}{1-(1-\gamma)L}\right] + \epsilon_t.$$

Rearranging,

$$(1-(1-\gamma)L)y_t = \beta\gamma x_t + (1-(1-\gamma)L)\epsilon_t.$$

Or,

$$y_t = (1-\gamma)y_{t-1} + \beta\gamma x_t + \epsilon_t - (1-\gamma)\epsilon_t.$$

This is an example of a dynamic stochastic equation. Notice that the expectations have induced dynamics. There are other ways as well to induce dynamics. One such way is to introduce costs of moving or changing. Hence the next model.

7.3.3 Partial Adjustment Model

- Suppose there exists a target outcome, y_t^* such that:

$$y_t^* = \beta x_t.$$

- Because of costly adjustment, y_t reacts as follows:

$$y_t = y_{t-1} + \gamma(y_t^* - y_{t-1}) + \epsilon_t, \quad \gamma \in [0,1].$$

Note here that we don't close the gap completely (as measured by γ) because doing so is costly. If $\gamma = 0$, then we get no adjustment. If $\gamma = 1$, we have full adjustment but this is costly. The γ parameter measures the "Speed of Adjustment."

- Substituting in for y_t^*, the model becomes:

$$y_t = (1-\gamma)y_{t-1} + \beta\gamma x_t + \epsilon_t.$$

Note that this dynamic stochastic difference equation is very similar to the one generated under rational expectations.

8 Week 8: 4 Mar - 8 Mar

8.1 A General Dynamic Model

- Suppose y_t^* is the equilibrium value of some random variable. (Output, employment, income, etc.)
- y_t is the actual observed value.
- The very general model is:

$$y_t^* = \beta_0 + \beta_1 x_t + u_t,$$

with x_t determined exogenously.

- Assume the agent solves the following minimization problem:

$$Min \quad E\left[\sum_{s=0}^{\infty} \alpha^s \Big(\underbrace{(y_{t+s} - y_{t+s}^*)^2}_{Deviation\ From\ Equilibrium} + \underbrace{\lambda(y_{t+s} - y_{t+s-1})^2}_{Cost\ of\ Moving} \Big)\right].$$

- Note that this is a stochastic problem because we have the expectation of a minimum operator. Thus, because it is a quadratic, the solution is also the solution to:

$$Min \quad \sum_{s=0}^{\infty} \alpha^s \Big((y_{t+s} - E_t[y_{t+s}^*])^2 + \lambda(y_{t+s} - y_{t+s-1})^2 \Big).$$

This is by the first order certainty equivalence. This is now a deterministic problem. Thus uncertainty does not effect the minimization decision. Note this only works (for some reason) for quadratics.

- The solution to this problem (taken on faith) is:

$$y_t = \mu y_{t-1} + (1-\mu)\left[(1-\alpha\mu)\sum_{i=1}^{\infty}(\alpha\mu)^i E_t[y_{t-i}^*]\right].$$

Where μ is the stable root of the following quadratic:

$$(\alpha\lambda)\mu^2 - (1 + \alpha\lambda + \lambda)\mu + \lambda = 0.$$

And y_t^* is given by:

$$y_t^* = \beta_0 + \beta_1 x_t + u_t.$$

- Some remarks about this solution.

- $\dfrac{\partial \mu}{\partial \lambda} > 0$. Thus since λ measures the adjustment costs, if μ is larger, then the future is more important (more weight is placed on the future values y^* in this case.)
- Note that $\sum_{i=1}^{\infty}(\alpha\mu)^i$ is a geometric series which equals $\frac{1}{1-\alpha\mu}$. Thus, this cancels with the $(1-\alpha\mu)$ outside the summation and the total weight overall time periods is one.
- This is very similar to a partial adjustment process.

- To make the solution operational, we need to eliminate y^*_{t+1}. We stated above that y^* evolved according to the process above as a function of x_t and u_t. x_t is a process determined outside the model so it is usually known. Assume, (only as an example) it takes the following form:

$$x_t = \rho x_{t-1} + \epsilon_t, \quad |\rho| < 1, \quad \epsilon_t \sim \ iid(0, \cdot), \quad u_t \sim \ iid(0, \cdot).$$

- Thus, if

$$y^*_t = \beta_0 + \beta_1 x_t + u_t.$$

Then

$$E_t[y^*_{t+i}] = \beta_0 + \beta_1 E_t[x_{t+i}].$$

- But consider the x_t process:

$$x_t = \rho x_{t-1} + \epsilon_t.$$

Taking expectations,

$$E_t[x_{t+i}] = \rho E_t[x_{t-1+i}].$$

Backward substitution:

$$E_t[x_{t+i}] = \rho^2 E_t[x_{t-2+i}].$$

$$E_t[x_{t+i}] = \rho^3 E_t[x_{t-3+i}].$$

$$\vdots$$

$$E_t[x_{t+i}] = \rho^i E_t[x_t] = \rho^i x_t.$$

- Substituting,

$$E_t[y^*_{t+i}] = \beta_0 + \beta_1 \rho^i x_t \quad for\ i > 0.$$

  Note case when $i = 0$, u_t is known, so:

$$y^*_t = \beta_0 + \beta_1 x_t + u_t \quad for\ i = 0.$$

- So now we will substitute these expressions for y^*_t back into our solution equation above:

$$y_t = \mu y_{t-1} + (1-\mu)\left[(1-\alpha\mu)\Big[\beta_0 \sum_{i=1}^{\infty}(\alpha\mu)^i + \beta_1 \sum_{i=1}^{\infty}(\alpha\mu\rho)^i x_t + u_t\Big]\right].$$

  Note again our geometric series: $\sum_{i=1}^{\infty}(\alpha\mu)^i = \dfrac{1}{1-\alpha\mu}$ and $\sum_{i=1}^{\infty}(\alpha\mu\rho)^i = \dfrac{1}{1-\alpha\mu\rho}$. Thus substituting these in and canceling,

$$y_t = \mu y_{t-1} + \beta_0(1-\mu) + \frac{\beta_1(1-\mu)(1-\alpha\mu)}{(1-\alpha\mu\rho)} x_t + u_t(1-\mu)(1-\alpha\mu).$$

  This equation along with $x_t = \rho x_{t-1} + \epsilon_t$ completes the model.

- At this point we have two options. We could either:
  - 1) Regression y on a constant, y_{-1}, and x.
  - 2) Treat the model seriously by estimating the deep parameters $(\beta_0, \beta_1, \alpha, \mu, \rho)$. Note we can't estimate all of these from the regression alone because the model is underidentified. (We would need an alternative measure of α).

- The choice between options 1 and 2 is exactly what Lucas was getting at in the Lucas Critique. Running the simple regression in option 1 does not take into account that the policy decision that you might make based on the regression results would influence the structural form of the system. (Particularly, the x_t process evolves according to the parameter ρ which also appears in the y_t equation above.) If the x_t process changes, then the relationship between y_t and x_t changes so the policy might not have the desired impact. So using option 2 is more difficult because the model is underidentified but it helps us get at the parameters that actually characterise the behavior of the agents in the model.

8.2 Error Correction

- Consider the very general model:

$$y_t = \alpha_1 y_{t-1} + \alpha_2 y_{t-2} + \cdots + \alpha_m y_{t-m} + \beta_0 x_t + \beta_1 x_{t-1} + \beta_2 x_{t-2} + \cdots + \beta_n x_{t-n} + \epsilon_t.$$
$$\epsilon_t \sim \ iid(0, \sigma^2).$$

- We could also rewrite the model as:

$$A(L)y_t = B(L)x_t + \epsilon_t.$$

With:

$$A(L) = 1 - \alpha_1 L - \alpha_2 L^2 - \cdots - \alpha_m L^m.$$

And,

$$B(L) = \beta_0 + \beta_1 L + \beta_2 L^2 + \cdots + \beta_n L^n.$$

- It is easy to see that the immediate (instantaneous) effect of a change in x on y is embodied in β_0. So β_0 could be thought of as the short run multiplier. In later periods the mulitplier process continues involving both the β's and the α's.

- Now consider the long run multiplier of x on y. An easy way to show this is to rewrite the model and ignore the time subscripts:

$$y = \alpha_1 y + \alpha_2 y + \cdots + \alpha_m y + \beta_0 x + \beta_1 x + \beta_2 x + \cdots + \beta_n x.$$
$$y(1 - \alpha_1 - \alpha_2 - \cdots - \alpha_m) = x(\beta_0 + \beta_1 + \beta_2 + \cdots + \beta_n).$$
$$y(A(1)) = x(B(1)).$$
$$\frac{y}{x} = \frac{B(1)}{A(1)}.$$

So the total (long run) effect of a change in x on y is equal to $\frac{B(1)}{A(1)}$.

8.3 More on the General Dynamic Model

- Consider again the general dynamic model and this time we'll include a constant term:

$$A(L)y_t = \alpha_0 + B(L)x_t + \epsilon_t, \quad \epsilon_t \sim \ iid(0, \sigma^2).$$

Where,

$$A(L) = 1 - \alpha_1 L - \alpha_2 L^2 - \cdots - \alpha_m L^m.$$

And,

$$B(L) = \beta_0 + \beta_1 L + \beta_2 L^2 + \cdots + \beta_n L^n.$$

- Ignoring the time subscripts and evalutating at $A(1)$ and $B(1)$, we obtain the long run relationship as before:

$$y = \frac{\alpha_0}{A(1)} + \frac{B(1)}{A(1)}x.$$

Where $\frac{B(1)}{A(1)}$ is the long run multiplier.

- The next step is to rewrite this model in Error Correction Form. First consider the $A(L)$ and $B(L)$ polynomials. Rewrite $A(L)$ as:

$$A(L) = (1-L)+(1-\alpha_1-\alpha_2-\cdots-\alpha_m)L-\alpha_2(L^2-L)-\alpha_3(L^3-L)-\cdots-\alpha_m(L^m-L).$$

Thus,

$$A(L) = (1-L) + A(1)L + \alpha_2(L-L^2) + \alpha_3(L-L^3) + \cdots + \alpha_m(L-L^m).$$

Or,

$$A(L) = (1-L) + A(1)L + \alpha_2(L-L^2) + \alpha_3((L-L^2) + (L^2-L^3)) + \ldots$$
$$+\alpha_m((L-L^2) + (L^2-L^3) + \cdots + (L^{m-1}-L^m)).$$

Multiplying everything by y_t,

$$A(L)y_t = (1-L)y_t + A(1)Ly_t + \alpha_2(L-L^2)y_t + \alpha_3((L-L^2)+(L^2-L^3))y_t + \ldots$$
$$+\alpha_m((L-L^2) + (L^2-L^3) + \cdots + (L^{m-1}-L^m))y_t.$$

$$A(L)y_t = \Delta y_t + A(1)y_{t-1} + (\alpha_2+\alpha_3+\cdots+\alpha_m)\Delta y_{t-1}$$
$$+(\alpha_3+\alpha_4+\cdots+\alpha_m)\Delta y_{t-2} + \cdots + (\alpha_m)\Delta y_{t-m+1}.$$

Let $\alpha_j^* = (\alpha_j + \alpha_{j+1} + \cdots + \alpha_m)$. Thus,

$$A(L)y_t = \Delta y_t + A(1)y_{t-1} + \alpha_2^*\Delta y_{t-1} + \alpha_3^*\Delta y_{t-2} + \cdots + \alpha_m^*\Delta y_{t-m+1}.$$

- Similarly for $B(L)$ we can rewrite it as,

$$B(L)x_t = B(1)x_{t-1} + \beta_0\Delta x_t - \beta_2^*\Delta x_{t-1} - \beta_3^*\Delta x_{t-2} - \cdots - \beta_n^*\Delta x_{t-n+1}.$$

With $\beta_j^* = (\beta_j + \beta_{j+1} + \cdots + \beta_n)$.

- So substituting in,

$$A(L)y_t = \alpha_0 + B(L)x_t + \epsilon_t.$$

$$\Delta y_t + A(1)y_{t-1} + \alpha_2^* \Delta y_{t-1} + \alpha_3^* \Delta y_{t-2} + \cdots + \alpha_m^* \delta y_{t-m+1} = \alpha_0 + B(1)x_{t-1} + \beta_0 \Delta x_t - \beta_2^* \Delta x_{t-1}$$
$$-\beta_3^* \Delta x_{t-2} - \cdots - \beta_n^* \Delta x_{t-n+1} + \epsilon_t.$$

  And rearranging,

$$\Delta y_t = \alpha_0 - A(1)y_{t-1} + B(1)x_{t-1} - \alpha_2^* \Delta y_{t-1} - \alpha_3^* \Delta y_{t-2} - \cdots - \alpha_m^* \Delta y_{t-m+1}$$
$$+\beta_0 \Delta x_t - \beta_2^* \Delta x_{t-1} - \beta_3^* \Delta x_{t-2} - \cdots - \beta_n^* \Delta x_{t-n+1} + \epsilon_t.$$

  And this, finally, is the "Error Correction Form."

- Again, we can determine the long run solution, ie, when y_t and x_t are stationary so the Δy and Δx terms are all 0. Thus, dropping time subscripts we have,

$$0 = \alpha_0 - A(1)y + B(1)x.$$

  Or,

$$y = \frac{\alpha_0}{A(1)} + \frac{B(1)}{A(1)}x$$

  as before.

- To see why this is called Error Correction form, rewrite the model as:

$$\Delta y_t = -A(1)\left[y_{t-1} \underbrace{- \frac{\alpha_0}{A(1)} - \frac{B(1)}{A(1)}x_{t-1}}_{Long\ Run\ Value} \right] + \underbrace{\text{the } \Delta x \text{ and } \Delta y \text{ terms}}_{Short\ Run\ Dynamics}.$$

  So we can see that if the value of y at period $t-1$ is different from the long run stable solution predicted by x at time $t-1$, then we have movement in Δy_t. For example if $y_{t-1} >$ the long run value, then because of the negative sign, Δy moves down.

8.4 Useful Test in Modeling

- Standard F test on Linear Restrictions.
- Parameter Stability Tests.
  - Suppose we have a situation where the model is a good fit, but the parameters have to be adjusted over time. Thus the model has instable parameters. Consider the maintained model:

$$y_t = \beta_{1,0} + \beta_{1,1}x_{1,t} + \beta_{1,2}x_{2,t} + \cdots + \beta_{1,K-1}x_{K-1,t} + \epsilon_t \quad for\ t = 1...T_1.$$

    And,

$$y_t = \beta_{2,0} + \beta_{2,1}x_{1,t} + \beta_{2,2}x_{2,t} + \cdots + \beta_{2,K-1}x_{K-1,t} + \epsilon_t \quad for\ t = T_1 + 1...T.$$

  - Thus, under H_0, the null hypothesis is that the parameters are robust, we have K restrictions of the form:

$$H_0 : \beta_{1,0} = \beta_{2,0}, \beta_{1,1} = \beta_{2,1}, \ldots, \beta_{1,K-1} = \beta_{2,K-1}.$$

  - An easy way to test this hypothesis is to define a dummy variable as follows:

$$D_t = \begin{cases} 0 & \text{if } t \leq T_1 \\ 1 & \text{if } t > T_1 \end{cases} \tag{17}$$

  - The maintained model can now be written,

$$y_t = \beta_{1,0} + \beta_{1,1}x_{1,t} + \beta_{1,2}x_{2,t} + \cdots + \beta_{1,K-1}x_{K-1,t}$$
$$+(\beta_{2,0} - \beta_{1,0})D_t + (\beta_{2,1} - \beta_{1,1})D_t x_{1,t} + \cdots + (\beta_{2,K-1} - \beta_{1,K-1})D_t x_{K-1,t} + \epsilon_t.$$

  - Under H_0, the coefficents on $D_t, D_t x_{1,t}, D_t x_{2,t}, \ldots, D_t x_{K-1,t}$ are all ZERO.
  - So the procedure is as follows: 1) Run regression on $t = 1...T$ and record RSS. 2) Run regression on $t = 1...T_1$ and record RSS_1. 1) Run regression on $t = T_1 + 1...T$ and record RSS_2. Under H_0,

$$F = \frac{(RSS - RSS_1 - RSS_2)/K}{(RSS_1 + RSS_2)/T - 2K} \sim F(K, T - 2K).$$

  - One final note: If $(T_1 + 1...T)$ has less than K observations, then you must use the following test statistic because RSS_2 is essentially 0:

$$F = \frac{(RSS - RSS_1)/T - T_1}{(RSS_1)/T_1 - K} \sim F(T - T_1, T_1 - K).$$

- Serial Correlation Test. Use Lagrange Multiplier Test.

- Common Factor Test.
  - Maintained model as follows:

$$y_t = \phi y_{t-1} + x_t'\beta - x_{t-1}'\delta + \epsilon_t.$$

  - Null hypothesis:

$$y_t = x_t'\beta + u_t, \quad u_t = \phi u_{t-1} + \epsilon_t, \quad \epsilon_t \sim \ iid(0, \sigma^2).$$

  - Note that under H_0, if we lag y_t, multiply through by ϕ and subtract the result from y_t, we get,

$$y_t - \phi y_{t-1} = x_t'\beta - x_{t-1}'\beta\phi + u_t - \phi u_{t-1}.$$

$$y_t = \phi y_{t-1} + x_t'\beta - x_{t-1}'\beta\phi + \epsilon_t.$$

  - Thus under H_0, $\delta = \beta\phi$. So if there are K, x variables then there are K restrictions of the form:

$$\frac{\delta_i}{\beta_i} = \phi \quad for\ i = 1...K.$$

  - Use the Likelihood ratio test to test the null.

8.5 Strategies in Time Series Modeling

- 1.) Take each variable and investigate if it has a unit root.
- 2.) Suppose that all variables do (which is usually the case in practice). Write down a long run relationship, ie:

$$y_t = \beta_0 + \sum_{k=1}^{K} \beta_k x_{kt} + \epsilon_t.$$

  Note here we exclude the short run dynamic terms (the delta terms). If the model is reasonable, there should exist a set of β parameters, $\hat{\beta}_k$, so that $y_t - \sum_{k=1}^{K} \hat{\beta}_k x_{kt}$ is stationary. If this is the case then $y_t, x_{1t}, x_{2t}, \ldots, x_{Kt}$ are said to be "Co-integrated" and $(1, -\beta_1, -\beta_2, \ldots, -\beta_K)$ is the "Co-Integrating Vector."

- 3.) Test for Co-Integration? Regress y on x's using OLS. Generate residuals, $\hat{\epsilon}_t$, and test these for stationarity via a Dicky-Fuller test. Note, however that we must use slightly different critical values (MacKinnon Tables) because we are estimating $\hat{\epsilon}$.
- More next week.

9 Week 9: 11 Mar - 15 Mar

9.1 More Strategies in Time Series Modeling

- We continue from last week with more strategies in time series modeling.
- 4.) If you can reject the null hypothesis of NO-cointegration, ie there is a long run relationship, then estimate a dynamic model such that:

  $$A(L)y_t = \alpha_0 + B(L)x_t + \epsilon_t.$$

  Or in Error Correction Form (ECF):

  $$\Delta y_t = \alpha_0 - A(1)y_{t-1} + B(1)x_{t-1} - \alpha_1^*\Delta y_{t-1} - \cdots - \alpha_m^*\Delta y_{t-m+1}$$
  $$+\beta_0^*\Delta x_t + \beta_1^*\Delta x_{t-1} + \cdots + \beta_n^*\Delta x_{t-n+1} + \epsilon_t.$$

  With:

  $$\beta_j^* = (\beta_j + \beta_{j+1} + \cdots + \beta_n).$$
  $$\alpha_j^* = (\alpha_j + \alpha_{j+1} + \cdots + \alpha_m).$$

  Note the ECF is a little more convenient for hypothesis testing and inference procedures. Note that if the coefficient on the level term, y_{t-1}, is $A(1) = 0$, then there is NO long run relationship.
  How many lags to include?

  - As a general rule, use 5 lags for quarterly data and 2 lags for annual data. However you have to use common sense when determining exactly how many because you want to leave yourself with some degrees of freedom in the end.

- 5.) Check errors to make sure they are NOT serially correlated. If there are, then add lags (especially on the dependent variable).
- 6.) Refine the model by removing highly insignificant Delta terms. Note: NEVER throw away any level terms before all the Delta terms have been ruled out. Also, don't just throw away ALL insignificant terms, only those that are highly insignificant. It's best to leave them in than to create omitted variable bias.
- 7.) Final step: Test for Parameter Stability. Parameters will become unstable either because the behavior of the agents in the model have changed (unlikely), or because the model is misspecified (more than likely). So why does misspecification generate parameter instability. This is best seen with the following example.

9.1.1 Model Misspecification Generates Parameter Instability

- Suppose the true model is the following:

$$y_t = \alpha y_{t-1} + (1-\alpha)x_t + u_t, \quad u_t \sim \; iid(0, \sigma_u^2), \quad \alpha \in (0,1).$$

Thus the short run multiplier is $(1-\alpha)$ and the long run mulitplier is found by dropping the time subscripts and the error term:

$$y = \alpha y + (1-\alpha)x.$$
$$y(1-\alpha) + (1-\alpha)x.$$
$$y = x.$$
$$\frac{y}{x} = 1 = \text{Long Run Multipler}.$$

- Suppose someone mistakenly estimates the following static model:

$$y_t = \gamma x_t + \epsilon_t.$$

- The usual OLS estimate of γ is:

$$\hat{\gamma} = \frac{\sum x_t y_t}{\sum x_t^2}.$$

- Suppose x_t is generated by the following process:

$$x_t = \rho x_{t-1} + \eta_t, \quad \eta_t \sim \; (0, \sigma_\eta^2), \quad |\rho| < 1.$$

From this simple $AR(1)$ process, we know that:

$$E[x_t \cdot x_{t-i}] = \frac{\rho^i \sigma_\eta^2}{1-\rho^2}.$$

- Now back to the true model:

$$y_t = \alpha y_{t-1} + (1-\alpha)x_t + u_t.$$

Backward substituting:

$$y_t = \alpha(\alpha y_{t-2} + (1-\alpha)x_{t-1} + u_{t-1}) + (1-\alpha)x_t + u_t.$$
$$y_t = \alpha^2 y_{t-2} + \alpha(1-\alpha)x_{t-1} + \alpha u_{t-1} + (1-\alpha)x_t + u_t.$$
$$y_t = \alpha^2 y_{t-2} + (1-\alpha)(x_t + \alpha x_{t-1}) + u_t + \alpha u_{t-1}.$$

And after infinitely substituting like we have here, we get the following:

$$y_t = (1-\alpha)(x_t + \alpha x_{t-1} + \alpha^2 x_{t-2} + \dots) + (u_t + \alpha u_{t-1} + \alpha^2 u_{t-2} + \dots).$$

- Thus,

$$E[x_t y_t] = (1-\alpha)E[\underbrace{x_t^2}_{\sigma^2/(1-\rho^2)} + \alpha \underbrace{x_t x_{t-1}}_{\rho\sigma^2/(1-\rho^2)} + \alpha^2 \underbrace{x_t x_{t-2}}_{\rho^2\sigma^2/(1-\rho^2)} + \dots).$$

Because x_t is uncorrelated with all the u_t's because u_s and η_t are independent for all s and t. Thus,

$$E[x_t y_t] = (1-\alpha)\frac{\sigma_\eta^2}{1-\rho^2}\underbrace{[1 + \alpha\rho + \alpha^2\rho^2 + \dots]}_{1/(1-\alpha\rho)}.$$

$$E[x_t y_t] = (1-\alpha)\frac{\sigma_\eta^2}{1-\rho^2} \cdot \frac{1}{1-\alpha\rho}.$$

$$= \frac{(1-\alpha)\sigma_\eta^2}{(1-\rho^2)(1-\alpha\rho)}.$$

- So by slutsky,

$$plim\hat{\gamma} = \frac{plim\frac{1}{T}\sum x_t y_t}{plim\frac{1}{T}\sum x_t^2}.$$

And by the ergodic theorem,

$$\frac{plim\frac{1}{T}\sum x_t y_t}{plim\frac{1}{T}\sum x_t^2} = \frac{E[x_t y_t]}{E[x_t^2]}.$$

Thus because $E[x_t^2] = \sigma_\eta^2/(1-\rho^2)$,

$$plim\hat{\gamma} = \frac{(1-\alpha)\sigma_\eta^2}{(1-\rho^2)(1-\alpha\rho)} \cdot \frac{1-\rho^2}{\sigma_\eta^2}.$$

Or,

$$plim\hat{\gamma} = \frac{(1-\alpha)}{(1-\alpha\rho)}.$$

- So given $0 < \rho < 1$, $(1-\alpha) < plim\hat{\gamma} < 1$, or:

$$\text{Short Run Multiplier} < plim\hat{\gamma} < \text{Long Run Multiplier.}$$

Thus if the x process changes, ie, ρ changes, then $plim\hat{\gamma}$ will change even if the parameter of the true model, α, does NOT change. Thus we get parameter instability from model misspecification.

9.2 Systems of Equations

- The Multivariate Regression. Consider a set of N regressions of the form:

$$y_t = X_t\beta + \epsilon_t.$$

Where y_t is a $Nx1$ vector, X_t is a $NxK = \sum_i K_i$ matrix of row vectors, β is a N vector of K vectors of coefficients and ϵ_t is a N vector. Thus we have:

$$\begin{bmatrix} y_{1t} \\ y_{2t} \\ \vdots \\ y_{Nt} \end{bmatrix} = \begin{bmatrix} x'_{1t} & 0 & \dots & 0 \\ 0 & x'_{2t} & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \dots & 0 & x'_{Nt} \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_K \end{bmatrix} + \begin{bmatrix} \epsilon_{1t} \\ \epsilon_{2t} \\ \vdots \\ \epsilon_{Nt} \end{bmatrix}. \tag{18}$$

- So the individual regressions are of the form:

$$y_{1t} = x'_{1t}\beta_1 + \epsilon_{1t}.$$

- Suppose that $E[\epsilon_t\epsilon'_t] = \Omega$. Thus we have the potential for correlations.
- We could estimate by OLS on each equation and get:

$$\hat{\beta} = (\sum_{t=1}^{T} X'_t X_t)^{-1} \sum_{t=1}^{T} X'_t y_t.$$

- However, in general, GLS is more efficient. First do OLS as above, generate $\hat{\epsilon}_t$, and then generate:

$$\hat{\Omega} = \sum_{t-1}^{T} \hat{\epsilon}_t\hat{\epsilon}'_t.$$

Then estimate:

$$\hat{\beta}_{GLS} = (\sum_{t=1}^{T} X'_t \hat{\Omega}^{-1} X_t)^{-1} \sum_{t=1}^{T} X'_t \hat{\Omega}^{-1} y_t.$$

This simplifies to OLS on each equation if either:

- 1) Ω is diagonal (no error correlations)
- 2) The regressors in each equation are the same.

- In this frame work, all x's are, at least, contemporaneously independent of ϵ. If NOT, we have a more general model.

- A More General Model.

$$\Gamma y_t = Bx_t + \epsilon_t.$$

Where the x_t's are, at least, contemporaneously independent. So we have N equations, N endogenous variables (y_t) and K exogenous variables (x_t). Thus Γ is NxN and B is NxK.

- We solve for y in terms of x_t or:

$$y_t = \Gamma^{-1}Bx_t + \Gamma^{-1}\epsilon_t.$$

Or letting $\Gamma^{-1}B = \Pi$ and $\Gamma^{-1}\epsilon_t = v_t$,

$$y_t = \Pi x_t + v_t.$$

This is often known as the "Data Generating Process."

9.3 More on Systems - Identification

- Consider again the structural form developed in the last section:

$$\Gamma y_t = Bx_t + \epsilon_t.$$

Or rewritten in reduced form:

$$y_t = \Gamma^{-1}Bx_t + \Gamma^{-1}\epsilon_t.$$
$$y_t = \Pi x_t + v_t.$$

This final form is refered to as the "Data Generating Process."

- It is clear therefore that $\Gamma^{-1}B = \Pi$ or that:

$$B = \Gamma\Pi.$$

Since B is NxK, we basically have NxK equations.

- If we normalize the diagonal elements of Γ to be 1, there are $(N-1)xN$ unknowns in Γ and NxK unknowns in B. Thus, we have more unknowns than equations and the system is NOT identified.
- Consider the following example.

9.3.1 Classic Supply and Demand Determination

- Consider the following two equations specifying supply and demand in an economy:

$$Demand : y_t = \alpha_0 - \alpha_1 p_t + u_t.$$

$$Supply : y_t = \beta_0 + \beta_1 p_t + v_t.$$

If we collect data points of price and output levels for this economy, we will just get points in a scatter that does not seem to follow either supply or demand. The problem is that both equations are functions of y and p so the points are equilibrium points for the whole system. Plotting these points will not let us find the supply and demand relationships individually.

- Thus we look to add another term to the supply function that is particular to supply and has no effect on demand. In the case of agricultural data, it can be argued that the weather conditions have a large impact on supply but virtually no impact on demand. Thus we write:

$$Supply : y_t = \beta_0 + \beta_1 p_t + v_t + W_{t-1}.$$

Where W is an index of the weather conditions in the prior period.

- Now if we plot (y, p) combinations over varying weather conditions we begin to plot out what looks like a demand curve. See notes for graph. [**G-9.1**] Note that if weather also impacted the demand equation, we would be back where we started. This assumption about the weather is an a priori asssumption.
- To estimate the supply equation we could include something in the demand equation (like income) that in general does not influence supply.

9.3.2 Conditions for Identification

- 1.) The conditions apply to one equation at a time.
- 2.) The Order Condition. A necessary condition is that the number of omitted y and x variables is greater than or equal to $N-1$. A linear restriction counts as a missing variable.

- 3.) The Rank Condition. A necessary and sufficient condition is that the matrix obtained from all the Γ and B coefficients in the other equations corresponding to the zeros in the equation concerned is of rank $N-1$.
- 4.) If the Rank Condition is passed and the number of zeros plus linear restrictions are greater than $N-1$, the equation is "over-identified." There are more restrictions than needed.
- 5.) An identity is always identified.

9.3.3 An Example of Identification

- Consider the following system:

$$y_{1t} = \gamma_{12}y_{2t} + \beta_{11}z_{1t} + \beta_{13}z_{3t} + \beta_{14}z_{4t} + u_{1t}.$$
$$y_{2t} = \gamma_{24}y_{4t} + \beta_{22}z_{2t} + \beta_{24}z_{4t} + u_{2t}.$$
$$y_{3t} = \gamma_{32}y_{2t} + \beta_{32}z_{2t} + \beta_{33}z_{3t} + u_{3t}.$$
$$y_{4t} = \gamma_{42}y_{2t} + \gamma_{43}y_{3t} + \beta_{41}z_{1t} + \beta_{42}z_{2t} + \beta_{44}z_{4t} + u_{4t}.$$

- To assess the level of identification, construct the following table:

y_{1t}	y_{2t}	y_{3t}	y_{4t}	z_{1t}	z_{2t}	z_{3t}	z_{4t}
1	$-\gamma_{12}$	0	0	$-\beta_{11}$	0	$-\beta_{13}$	$-\beta_{14}$
0	1	0	$-\gamma_{24}$	0	$-\beta_{22}$	0	$-\beta_{24}$
0	$-\gamma_{32}$	1	0	0	$-\beta_{32}$	$-\beta_{33}$	0
0	$-\gamma_{42}$	$-\gamma_{43}$	1	$-\beta_{41}$	$-\beta_{42}$	0	$-\beta_{44}$

Note that all coefficients have the signs as if they had been moved to the left hand side of the equation.

- Now we will use this table to analyze the identification of each equation. Note that $N-1=3$.
- Equation 1. Looking across the first row, there are 3 zeros. Thus,

Number of zeros $= 3 = N-1 = 3$. Thus, Order Condition Passed.

For the rank condition, we consider the rank of the matrix formed by taking the coefficients of the equations in the columns that contain zeros in equation 1. We seek:

$$rank \begin{bmatrix} 0 & -\gamma_{24} & -\beta_{22} \\ 1 & 0 & -\beta_{32} \\ -\gamma_{43} & 1 & -\beta_{42} \end{bmatrix}. \tag{19}$$

This rank will be equal to 3 (the min of number of columns and rows) iff the determinant of this matrix is NOT 0. By inspection, the determinant is $\neq 0$ and therefore, Rank = $N-1 = 3 =$ full rank. Thus, Rank Condition Passed. Thus y_{1t} is JUST IDENTIFIED.

- Equation 2. Looking across the second row, there are 4 zeros. Thus,

$$\text{Number of zeros} = 4 > N-1 = 3. \text{ Thus, Order Condition Passed.}$$

For the rank condition, we consider the rank of the matrix formed by taking the coefficients of the equations in the columns that contain zeros in equation 2. We seek:

$$rank \begin{bmatrix} 1 & 0 & -\beta_{11} & -\beta_{13} \\ 0 & 1 & 0 & -\beta_{33} \\ 0 & -\gamma_{43} & -\beta_{41} & 0 \end{bmatrix} \tag{20}$$

This rank will be equal to 3 (the min of number of columns and rows) iff the determinant of any square 3x3 matrix within this matrix is NOT 0. by inspection,

The determinant $\neq 0, \rightarrow Rank = N-1 = 3 =$ full rank. Thus, Rank Condition Passed.

Thus y_{2t} is OVER IDENTIFIED because of the 4 zeros.

- Equation 3. The same as 2. Overidentified.
- Equation 4. Order Condition: note the number of zeros in the fourth row is 2 so since $2 < N-1 = 3$, the order condition is not satisfied and equation 4 is NOT IDENTIFIED. In order to estimate y_{4t}, we would need another restriction. We cannot consistently estimate the fourth equation, but the other 3 are fine.

9.4 Estimation in Systems of Equations

- Consider again the structural form:

$$\Gamma y_t = Bx_t + \epsilon_t.$$

- Consider estimating the following equation:

$$y_{1t} = Y_{1t}'\gamma_1 + X_{1t}'\beta_1 + \epsilon_{1t}.$$

Where Y_{1t}' is a row vector of included y variables and X_{1t}' is a row vector of included x variables. Note we MUST have some excluded variables because otherwise the equation would not be identified.

- In matrix form we can write this as:

$$y_1 = Y_1\gamma_1 + X_1\beta_1 + \epsilon_1.$$

Where Y_1 is $(Tx(N_1 - 1))$. X_1 is TxK_1. N_1 is the number of y variables included in equation 1. K_1 is the number of x variables included in equation 1. So $N - N_1$ and $K - K_1$ is the number of excluded (y and x) variables (respectively).

- Let $X_1^\dagger$ be a $(Tx(K - K_1))$ matrix of excluded x variables. These will be used as instruments for Y_1. Are there enough? Y_1 needs at least $N_1 - 1$ instruments. So enough in $X_1^\dagger$ implies that:

$$K - K_1 \geq N_1 - 1.$$

Or,

$$(K - K_1) + (N - N_1) \geq (N_1 - 1) + (N - N_1).$$
$$\underbrace{(K - K_1) + (N - N_1)}_{Num.\ of\ Zeros} \geq N - 1.$$

Thus if the number of zeros is greater than $N - 1$, we will have enough instruments. ie, as long as the equation is identified.

10 Week 10: 18 Mar - 22 Mar

10.1 More on Simultaneous Systems of Equations

- Consider estimating the following equation:

$$y_{1t} = Y_{1t}'\gamma_1 + X_{1t}'\beta_1 + \epsilon_{1t}.$$

Where Y_{1t}' is a row vector of included y variables and X_{1t}' is a row vector of included x variables.

- In matrix form we can write this as:

$$y_1 = Y_1\gamma_1 + X_1\beta_1 + \epsilon_1.$$

- Instrumental variables techique. Suppose,

$$\begin{aligned} Z_1 &= [Y_1|X_1], \\ X_1 &= [X_1|X_1^{\dagger}], \\ \delta_1 &= \begin{bmatrix} \gamma_1 \\ \beta_1 \end{bmatrix}. \end{aligned} \tag{21}$$

Thus,

$$y_1 = Z_1\delta_1 + \epsilon_1.$$

- Thus the instrumental variables estimator is:

$$\hat{\delta}_1^{IV} = (Z_1'X(X'X)^{-1}X'Z_1)^{-1}(Z_1'X(X'X)^{-1}X'y_1).$$

And,

$$Var(\hat{\delta}_1^{IV}) = s_{IV}^2[Z_1'X(X'X)^{-1}X'Z_1]^{-1}.$$

- This derivation of $\hat{\delta}_1^{IV}$ was originally referred to as 2 stage least squares ($2SLS$). Take Z_1 and regress on exogenous variables, X. Take fitted values:

$$\hat{Z}_1 = X(X'X)^{-1}X'Z_1.$$

Then do OLS of y_1 on $\hat{Z}_1$. ie, compute:

$$\hat{\delta} = (\hat{Z}_1'\hat{Z}_1)^{-1}\hat{Z}_1'y_1.$$

Substituting in $\hat{Z}_1$, we see that:

$$\hat{\delta} = \hat{\delta}_1^{IV}.$$

So we can get at our instrumental variable estimator, δ^{IV}, via a 2 stage least squares process.

- One important note: If the equation is JUST identified, then $(Z_1'X)$ is square and thus invertible. We can then rearrange the terms of δ^{IV} above by taking the individual inverses. Thus,

$$\hat{\delta}^{IV} = (X'Z_1)^{-1}(X'X)(Z_1'X)^{-1}(Z_1'X(X'X)^{-1}X'y_1).$$

Which simplifes to:

$$\hat{\delta}^{IV} = (X'Z_1)^{-1}X'y_1.$$

This method is called "Indirect Least Squares." Note it only works if the equation is exactly or just identified.

10.2 Dynamic Models

- Consider the following structural form of a model:

$$A_0 y_t = A_1 y_{t-1} + A_2 y_{t-2} + \cdots + A_p y_{t-p} + B_0 x_t + B_1 x_{t-1} + \cdots + B_s x_{t-s} + \epsilon_t.$$

Where y_t is an N vector and the lagged y's and x's are contemporaneously independent.

- Solving for y_t, we get the reduced form:

$$y_t = A_0^{-1}A_1 y_{t-1} + A_0^{-1}A_2 y_{t-2} + \cdots + A_0^{-1}A_p y_{t-p} + A_0^{-1}B_0 x_t + A_0^{-1}B_1 x_{t-1} + \cdots + A_0^{-1}B_s x_{t-s} + A_0^{-1}\epsilon_t.$$

- We could also take the structural form and write it in lag form:

$$A(L)y_t = B(L)x_t + \epsilon_t.$$

Where $A(L)$ and $B(L)$ are matrices of lagged polynomials. For example:

$$A(L) = A_0 + A_1 L - A_2 L^2 - \cdots - A_P L^P.$$

- Note the rule for inversion:

$$A(L)^{-1} = \frac{A^*(L)}{|A(L)|}.$$

Where A^* is the adjoint of A, or the transposed matrix of cofactors and $|A|$ is the determinant of A.

- For example, suppose:

$$A(L) = \begin{bmatrix} a_{01} - a_{11}L & a_{12}L \\ a_{21}L + a_{22}L^2 & a_{23}L \end{bmatrix}. \tag{22}$$

Then,

$$A^*(L) = \begin{bmatrix} a_{23}L & -a_{12}L \\ -a_{21}L - a_{22}L^2 & a_{01} - a_{11}L \end{bmatrix}. \tag{23}$$

And,

$$|A(L)| = a_{01}a_{23}L - a_{11}a_{23}L^2 - a_{12}a_{21}L^2 - a_{12}a_{22}L^3.$$

- So now go back to the lag form we were considering:

$$A(L)y_t = B(L)x_t + \epsilon_t.$$

Solve for y_t,

$$y_t = A(L)^{-1}B(L)x_t + A(L)^{-1}\epsilon_t.$$

Apply inverse rule:

$$y_t = \frac{A^*(L)}{|A(L)|}B(L)x_t + \frac{A^*(L)}{|A(L)|}\epsilon_t.$$

Moving over the determinant term:

$$|A(L)|y_t = A^*(L)B(L)x_t + A^*(L)\epsilon_t.$$

This is known as the "Autoregressive Final Form." In this form, each y_t can be written as lags only of itself and lagged x's. Furthermore, the lag structure is the same in every equation.

10.2.1 Example of a Macro Model

- Consider the following macro model:

$$c_t = \beta_{11} + \gamma_{13}y_t + \gamma_{14}y_{t-1} + \epsilon_{1t}.$$
$$i_t = \beta_{21} + \beta_{22}i_{t-1} + \epsilon_{2t}.$$
$$y_t = c_t + i_t.$$

So the first equation is a consumption function which depends on current and lagged income. The second equation just relates current investment to prior investment. And the third equation is an identity.

- We could check the identification of this system using the method developed last week.
- Eliminating c_t from first equation using the 3rd, we have a system of 2 equations:

$$y_t = \beta_{11} + \gamma_{13} y_t + \gamma_{14} y_{t-1} + i_t + \epsilon_{1t}.$$
$$i_t = \beta_{21} + \beta_{22} i_{t-1} + \epsilon_{2t}.$$

Moving all endogenous lags to the left and rewriting using the lag operator:

$$y_t - \gamma_{13} y_t - \gamma_{14} L y_t - i_t = \beta_{11} + \epsilon_{1t}.$$
$$i_t - \beta_{22} L i_t = \beta_{21} + \epsilon_{2t}.$$

Thus, factoring out:

$$y_t(1 - \gamma_{13} - \gamma_{14} L) - i_t = \beta_{11} + \epsilon_{1t}.$$
$$i_t(1 - \beta_{22} L) = \beta_{21} + \epsilon_{2t}.$$

And writing in matrix form:

$$\underbrace{\begin{bmatrix} 1 - \gamma_{13} - \gamma_{14} L & -1 \\ 0 & 1 - \beta_{22} L \end{bmatrix}}_{A} \begin{bmatrix} y_t \\ i_t \end{bmatrix} = \begin{bmatrix} \beta_{11} \\ \beta_{21} \end{bmatrix} + \begin{bmatrix} \epsilon_{1t} \\ \epsilon_{2t} \end{bmatrix}. \tag{24}$$

- We could then take the A matrix and compute A^* and $|A|$ as before. Thus we would get the Autoregressive Final Form as:

$$|A| \begin{bmatrix} y_t \\ i_t \end{bmatrix} = A^* \begin{bmatrix} \beta_{11} \\ \beta_{21} \end{bmatrix} + A^* \begin{bmatrix} \epsilon_{1t} \\ \epsilon_{2t} \end{bmatrix}. \tag{25}$$

- We could also examine the stability of the system. Taking the determinant of A, we have:

$$|A| = (1 - \gamma_{13}) - [\beta_{22}(1 - \gamma_{13}) + \gamma_{14}]L + \gamma_{14}\beta_{22} L^2.$$

Thus we examine the roots of the following polynomial:

$$(1 - \gamma_{13})\lambda^2 - [\beta_{22}(1 - \gamma_{13}) + \gamma_{14}]\lambda + \gamma_{14}\beta_{22}.$$

If the roots are in absolute value less than one, then the system is STABLE.

10.3 The Last Topic: The Econometrics of Consumption

- In this section we will consider the life cycle model of consumption. Assume there are perfect capital markets, a constant real interest rate, r, and our consumer is infinitely lived.
- At time t, our consumer has assets, A_t, and faces a random labor income stream: $\{y_t, y_{t+1}, y_{t+2}, \ldots, y_{t+s}\}$. Of course only y_t is known at time t.
- Assume income and consumption "occurs" (received and spent) at the start of the period. Interest is paid out at the end of the period.
- Budget Constraint:

$$A_{t+s+1} = (1+r)\Big[A_{t+s} + Y_{t+s} - C_{t+s}\Big], \qquad s = 0, 1, 2, \ldots.$$

- The individual's utility is a simple additive form and then the consumer maximizes:

$$max \ \sum_{s=0}^{\infty} \rho^s E_t\Big[u(C_{t+s})\Big].$$

  Where ρ is a discount factor reflecting the rate of time preference.

- Substituting out using the budget constraint:

$$max \ \sum_{s=0}^{\infty} \rho^s E_t\left[u\Big(A_{t+s} + y_{t+s} + \frac{A_{t+s+1}}{1+r}\Big)\right].$$

- The first order condition with respect to A_{t+s}. Note that in the previous time period, A_{t+s} occurs in the last term. Thus,

$$\frac{\partial}{\partial A_{t+s}} = \rho^s E_t[u'(C_{t+s})] - \frac{\rho^{s-1}}{1+r} E_t[u'(C_{t+s-1})] \quad for \ s = 0 \ldots \infty.$$

- Evaluating at s=1,

$$\rho E_t[u'(C_{t+1})] = \frac{1}{1+r} E_t[u'(C_t)].$$

  Note the expectations can be removed:

$$\rho E_t[u'(C_{t+1})] = \frac{1}{1+r} u'(C_t).$$

  And thus,

$$u'(C_t) = \rho(1+r)E_t[u'(C_{t+1})].$$

Letting $\rho(1+r) = 1$, we have:

$$u'(C_t) = E_t[u'(C_{t+1})].$$

Removing expectations,

$$u'(C_t) = u'(C_{t+1}) - \nu_{t+1}.$$

$$u'(C_{t+1}) = u'(C_t) + \nu_{t+1}.$$

Where $\nu_{t+1} = u'(C_{t+1}) - E_t[u'(C_{t+1})]$, the innovation in the marginal utility of consumption.

- Now investigate a Quadratic Utility model to see how consumption responds to innovations in y. In data analysis, we see a tendency for consumption to be smoother than income. Given our derivation above, we have:

$$C_t = E_t[C_{t+1}].$$

And in period $t+s$,

$$C_{t+s} = E_{t+s}[C_{t+s+1}].$$

Taking expectations,

$$E_t[C_{t+s}] = E_t\Big[E_{t+s}[C_{t+s+1}]\Big].$$

By the law of iterative expectations,

$$E_t[C_{t+s}] = E_t[C_{t+s+1}].$$

Because all the information from time t to time s is unknown at time t. Thus using backward iteration, we have:

$$E_t[C_{t+s}] = C_t,$$

or all future consumption is expected to be the same as today's level of consumption.

- We can consider the budget constraint:

$$A_{t+s+1} = (1+r)\Big[A_{t+s} + Y_{t+s} - C_{t+s}\Big], \qquad s = 0, 1, 2, \ldots.$$

And rewrite it in present value form:

$$\sum_{s=0}^{\infty} \frac{C_{t+s}}{(1+r)^s} = A_t + \sum_{s=0}^{\infty} \frac{y_{t+s}}{(1+r)^s}.$$

Taking expectations:

$$\sum_{s=0}^{\infty} \frac{E_t[C_{t+s}]}{(1+r)^s} = E_t[A_t] + \sum_{s=0}^{\infty} \frac{E_t[y_{t+s}]}{(1+r)^s}.$$

Note that $E_t[C_{t+s}] = C_t$ and $E_t[A_t] = A_t$, thus:

$$C_t \sum_{s=0}^{\infty} \frac{1}{(1+r)^s} = A_t + \sum_{s=0}^{\infty} \frac{E_t[y_{t+s}]}{(1+r)^s}.$$

Now on the left we just have an infinite geometric series which sums to $\dfrac{1}{1 - 1/(1+r)}$. Thus,

$$C_t \frac{1}{1 - 1/(1+r)} = A_t + \sum_{s=0}^{\infty} \frac{E_t[y_{t+s}]}{(1+r)^s}.$$

$$C_t = \frac{r}{1+r} A_t + \frac{r}{1+r} \sum_{s=0}^{\infty} \frac{E_t[y_{t+s}]}{(1+r)^s}.$$

So this expression is basically the permanent income hypothesis. Consumption is equal to interest income (discounted by $1+r$ because it is paid out at the end of the period) plus the expected future value of labor income. So basically you consume a proportion $\dfrac{r}{1+r}$ of your total wealth.

- Now lag one period:

$$C_{t-1} = \frac{r}{1+r} A_{t-1} + \frac{r}{1+r} \sum_{s=0}^{\infty} \frac{E_{t-1}[y_{t+s-1}]}{(1+r)^s}.$$

Multiply by $(1+r)$,

$$(1+r)C_{t-1} = rA_{t-1} + r \sum_{s=0}^{\infty} \frac{E_{t-1}[y_{t+s-1}]}{(1+r)^s}.$$

$$C_{t-1} = -rC_{t-1} + rA_{t-1} + r\sum_{s=0}^{\infty}\frac{E_{t-1}[y_{t+s-1}]}{(1+r)^s}.$$

Now pull out the first term ($s = 0$) out of the summation and rewrite the terms inside of the summation (and NOT it's limits):

$$C_{t-1} = -rC_{t-1} + rA_{t-1} + ry_{t-1} + \frac{r}{1+r}\sum_{s=0}^{\infty}\frac{E_{t-1}[y_{t+s}]}{(1+r)^s}.$$

$$C_{t-1} = \underbrace{r(-C_{t-1} + A_{t-1} + y_{t-1})}_{(rA_t)/(1+r)} + \frac{r}{1+r}\sum_{s=0}^{\infty}\frac{E_{t-1}[y_{t+s}]}{(1+r)^s}.$$

Note the substitution comes from the budget constraint directly. Note that the discounting of the summation by $1+r$ is rather mysterious. It should drop out when you multiply through by $1+r$. The reason it's still there probably has something to do with payments being made at the end of the period (and it makes the algebra work out for the next step). Thus subtracting C_{t-1} from C_t:

$$C_t - C_{t-1} = \Delta C_t = \frac{r}{1+r}A_t + \frac{r}{1+r}\sum_{s=0}^{\infty}\frac{E_t[y_{t+s}]}{(1+r)^s} - (rA_t)/(1+r) - \frac{r}{1+r}\sum_{s=0}^{\infty}\frac{E_{t-1}[y_{t+s}]}{(1+r)^s}.$$

Or,

$$\Delta C_t = \frac{r}{1+r}\sum_{s=0}^{\infty}\frac{(E_t - E_{t-1})[y_{t+s}]}{(1+r)^s}.$$

Thus the only thing that changes your level of consumption is changes in views about your future levels of income.

- To make the equation workable, we need to know what the process is that is driving y_t. Assume y_t is a stationary ARMA with mean μ such that:

$$Z_t = y_t - \mu$$

satisfies,

$$A(L)Z_t = B(L)\epsilon_t.$$

Where $A(L)$ and $B(L)$ are lag polynomials as usual. Thus,

$$Z_t = A(L)^{-1}B(L)\epsilon_t = \Gamma(L)\epsilon_t.$$

Thus,

$$y_t = \mu + \Gamma(L)\epsilon_t.$$

- Writing y_t out in expanded form:

$$y_t = \mu + \epsilon_{t+s} + \gamma_1 \epsilon_{t+s-1} + \gamma_2 \epsilon_{t+s-2} + \cdots + \gamma_{s-1}\epsilon_{t+1} + \gamma_s \epsilon_t + \gamma_{s+1}\epsilon_{t-1} + \dots.$$

- Thus at time t, only those ϵ's at period t or earlier will be known and the rest will be zero in expectation. Thus,

$$E_t[y_{t+s}] = \mu + \gamma_s \epsilon_t + \gamma_{s+1}\epsilon_{t-1} + \dots.$$

  And by the same logic,

$$E_{t-1}[y_{t+s}] = \mu + \gamma_{s+1}\epsilon_{t-1} + \dots.$$

- Subtracting,

$$E_t[y_{t+s}] - E_{t-1}[y_{t-s}] = (E_t - E_{t-1})[y_{t+s}] = \gamma_s \epsilon_t.$$

- We can now substitute this into our equation above for ΔC_t. Thus,

$$\Delta C_t = \frac{r}{1+r} \sum_{s=0}^{\infty} \frac{(E_t - E_{t-1})[y_{t+s}]}{(1+r)^s}.$$

$$\Delta C_t = \frac{r}{1+r} \sum_{s=0}^{\infty} \frac{\gamma_s \epsilon_t}{(1+r)^s}.$$

  Which can be rewritten,

$$\Delta C_t = \frac{r}{1+r} \Gamma\Big(\frac{1}{1+r}\Big)\epsilon_t.$$

  Or substituting back in $A(L)$ and $B(L)$,

$$\Delta C_t = \frac{r}{1+r} A\Big(\frac{1}{1+r}\Big)^{-1} B\Big(\frac{1}{1+r}\Big)\epsilon_t.$$

  Or finally,

$$\Delta C_t = \frac{r}{1+r} \frac{\sum \beta_s/(1+r)^s}{\sum \alpha_s/(1+r)^s}\epsilon_t.$$

  Thus changes in consumption are a proportional result of the shocks to income, ϵ_t.

- Done.