Economics 721: Econometrics *

Matthew Chesnes

Updated: December 25, 2005

*These are Matthew Chesnes' notes from a course taught by Ingmar Prucha.

1 Lecture 1: September 1, 2005

1.1 Course Outline

- In general, we will be relaxing the assumption that $E[uu'] = \sigma^2 I_T$ and instead write:

$$E[uu'] = \sigma^2 \Omega.$$

  This results in the BLUE Generalized Least Squares estimator (GLS) of:

$$\hat{\beta}_{GLS} = (X'\Omega^{-1}X)^{-1}X'\Omega^{-1}y,$$

  or the Feasible GLS estimator of:

$$\hat{\beta}_{FGLS} = (X'\hat{\Omega}^{-1}X)^{-1}X'\hat{\Omega}^{-1}y,$$

- We will consider various forms of the Ω matrix. Even with heterskadasticity, the OLS estimator of β will still be unbiased since all we required for unbiasedness was $E[X'u] = 0$. However the variance/covariance matrix of β will be different.

- We also might have serially correlated errors, seemingly unrelated regressions where there is correlations between the errors, $AR(p)$ and $ARMA(p, q)$ models, and finally we will address panel data of the form y_{ti}. t indexes time and i indexes a household.

1.2 Review of Lag Operators

- Consider the AR(1) process:

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

  Thus,

$$y_t = \frac{1}{1 - \alpha L}\epsilon_t = \sum_{i=0}^{\infty}(\alpha L)^i \epsilon_t.$$

  Or,

$$y_t = \sum_{i=0}^{\infty} \alpha^i \epsilon_{t-i} = lim_{n \to \infty} \sum_{i=0}^{n} \alpha^i \epsilon_{t-i}.$$

  But if ϵ_t is a random variable, we have to consider realizations of $\epsilon(\omega)$. If the limit converges in quadratic mean for almost all ω, then we say the random variable also converges.

1.3 Review of Difference Equations

- Consider an equation of the form:

$$y_t = a_1 y_{t-1} + a_2 y_{t-2} + \cdots + a_p y_{t-p} + s_t.$$

Or,

$$y_t - a_1 y_{t-1} - \cdots - a_p y_{t-p} = s_t.$$

- Solve by adding the particular solution to the homogenous solution. Consider the homogenous solution:

  $$y_t - a_1 y_{t-1} - \cdots - a_p y_{t-p} = 0.$$

  The solution is of the form: $y_t = c\lambda^t$, $\lambda \neq 0$. Substitute this in assuming $c = 0$:

  $$\lambda^t - a_1 \lambda^{t-1} - \cdots - a_p \lambda^{t-p} = 0.$$

  Divide by λ^{t-p}:

  $$\lambda^p - a_1 \lambda^{p-1} - \cdots - a_p = 0.$$

  And this is the characteristic polynomial which in general will have p roots. Thus the homogenous part of our solution is:

  $$y_t = c_1 \lambda_1^t + \cdots + c_p \lambda_p^t.$$

- See discussion class notes for more. You may use the steady state solution for the particular solution. If you have a differential equation, try $y(t) = ce^{\lambda t}$ as your guess.

2 Lecture 2: September 6, 2005

2.1 Review of Classical Linear Regression Model

- Consider the model $y = X\beta + u$, and our OLS estimator:

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

Then if $E[u|X] = 0$,

$$E[\hat{\beta}] = \beta + E_{X,u}[(X'X)^{-1}X'u] = \beta + E_X E[(X'X)^{-1}X'u|X] = \beta + E_X[(X'X)^{-1}E[u|X]] = \beta.$$

- So if $E[u|X] = 0$, then we get an unbiased estimator.

- Another classical assumption is that $E[uu'|X] = \sigma^2 I$, so there is no autocorrelation or heteroskedasticity.

- What about consistency. Make the following two assumptions:

$$1)\ \ plim\frac{1}{T}X'u = 0,$$

$$2)\ \ plim\frac{1}{T}X'X = Q,\ finite.$$

Then:

$$plim(\hat{\beta}) = \beta + plim(\frac{X'X}{T})^{-1}\frac{X'u}{T} = \beta + \underbrace{plim(\frac{X'X}{T})^{-1}}_{<\infty} * \underbrace{plim\frac{X'u}{T}}_{0} = \beta.$$

So with these two conditions, we get consistency.

2.2 Generalized Linear Regression Model

- Consider the following new set of assumptions:

  - (B.1) $E[u] = 0$.
  - (B.2) $E[uu'] = \sigma^2\Omega$, such that Ω is positive definite and known.
  - (B.3) X has full column rank and is non-stochastic.

- So while the error still has mean zero, we allow for both serial correlation and heteroskedasticity of the error term.

- Then our generalized least squares (GLS) estimator is:

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

- Note that since Ω is positive definite, so is Ω^{-1}. Thus we can decompose this into:

$$\Omega^{-1} = p'p,$$

where p is nonsingular. This is always possible but p may not be unique.

- Thus,

$$\begin{aligned} X'\Omega^{-1}X &= X'p'pX \\ \alpha'X'\Omega^{-1}X\alpha &= \alpha'X'p'\underbrace{pX\alpha}_{z} \\ &= z'z \\ &= \sum_i z_i^2 > 0 \text{ if } z \neq 0 \end{aligned}$$

But since $z = pX\alpha$ and X has full column rank and p is nonsingular, then $z = pX\alpha > 0$ as long as $\alpha \neq 0$. Thus we can invert $X'\Omega^{-1}X$ since it is nonsingular and $\hat{\beta}_{GLS}$ is well defined.

- Now consider the model:

$$y = X\beta + u,$$

where $E[uu'] = \sigma^2\Omega$. Note we condition everything still on X but we drop it for notation. Consider premultiplying by p:

$$py = pX\beta + pu,$$

or,

$$y_* = X_*\beta + u_*.$$

- Thus $E[u_*] = pE[u] = 0$ and:

$$\begin{aligned} E[u_*u_*'] &= pE[uu']p' \\ &= p\sigma^2\Omega p' \\ &= \sigma^2 p(p'p)^{-1}p' \\ &= \sigma^2 pp^{-1}p'^{-1}p' = \sigma^2 I. \end{aligned}$$

- Thus,

$$\begin{aligned} \tilde{\beta} = \hat{\beta}_{GLS} &= (X_*'X_*)^{-1}X_*'y_* \\ &= (X'p'pX)^{-1}X'p'py \\ &= (X'\Omega^{-1}X)^{-1}X'\Omega^{-1}y, \end{aligned}$$

which is the GLS estimator!

- If we have the special case there $\Omega = diag(\sigma_1^2, \ldots, \sigma_T^2)$ with zeros on the off diagonal, then Py is simply the values of y divided by the standard deviation of that observation (σ_t) as is Pu. So when we have heterskedasticity, we can simply scale the data by their variances as it changes over the sample.

- **Proposition** $E[\hat{\beta}_{GLS}] = \beta$ and:

$$VC(\hat{\beta}_{GLS}) = \sigma^2(X'\Omega^{-1}X)^{-1} = \sigma^2(X_*'X_*)^{-1},$$

  with $X_* = pX$.

- **Proposition** The Aitkem Theorem. Given B1, B2, and B3, $\hat{\beta}_{GLS}$ is BLUE. This is a generalization of the Gauss Markov theorem. Start with $\tilde{\beta} = Cy$ where $C = (X'\Omega^{-1}X')^{-1}X'\Omega^{-1} + D$. Then restrict to unbiased estimators which yields $DX = 0$. Finally compare the Variance/Covariance of $\hat{\beta}_{GLS}$ to the Var/Cov of $\tilde{\beta}$.

- **Proposition** $\hat{\beta}_{GLS} = arg\ min_b\ (y-Xb)'\Omega^{-1}\underbrace{(y-Xb)}_{u}$. We can simply differentiate this to solve for the GLS estimator. We minimize the weighted sum of squared residuals. This is also equal to:

$$\hat{\beta}_{GLS} = arg\ min \sum_{t=1}^{T}\sum_{s=1}^{T} \tilde{u}_t\tilde{u}_s\omega^{ts}.$$

- It makes sense that we should weight observations differently. If two particular points go right through the "true" regression line, we might as well just use those two points and throw out the rest. Thus we weight observations based on their relative variance.

- **Proposition** The estimator for σ^2. Consider the transformed model:

$$\hat{u}_* = y_* - X_*\hat{\beta}_{GLS} = p(y - X\hat{\beta}_{GLS}) = p\tilde{u}.$$

  Thus,

$$\tilde{\sigma}^2 = \frac{1}{T-K}\hat{u}_*'\hat{u}_* = \frac{1}{T-K}\tilde{u}'p'p\tilde{u} = \frac{1}{T-K}\tilde{u}'\Omega^{-1}\tilde{u}.$$

- **Proposition** Hypothesis testing. Given $u \sim N(0, \sigma^2\Omega)$, we simply operate with the transformed model $py = pX\beta + pu$. All the t and F statistics are the same as under the classical setting.

- For prediction, things change a bit. Consider a model:

$$y = X\beta + u,\ t = 1...T.$$

  And future values:

$$y_F = X_F\beta + u_F,\ t = T+1...S.$$

  We want to predict y_F. Thus:

$$y_F^P = X_F\tilde{\beta} + E[u_F|u].$$

In the classical model this last term was just zero, but here we have to include it due to the possibility of autocorrelation. Thus, consider the variance/covariance matrix:

$$E\left[\begin{bmatrix} u \\ u_F \end{bmatrix}\begin{bmatrix} u & u_F \end{bmatrix}\right] = \begin{bmatrix} E[uu'] & E[uu_F'] \\ E[u_F u'] & E[u_F u_F'] \end{bmatrix} = \sigma^2 \begin{bmatrix} \Omega & \Omega_{12} \\ \Omega_{21} & \Omega_F \end{bmatrix}.$$

So if the u's are normal, then $E[u_F|u] = \Omega_{21}\Omega^{-1}u$, and thus:

$$y_F^p = X_F\tilde{\beta} + \Omega_{21}\Omega^{-1}\tilde{u}.$$

- **Theorem** Under normality, the GLS estimator is the maximum likelihood estimator.
- **Proposition** y_F^p above is BLUE.

3 Lecture 3: September 8, 2005

3.1 More on GLS

- When Ω is known, our (True) GLS estimator is:

$$\tilde{\beta} = (X'\Omega^{-1}X)^{-1}X'\Omega^{-1}y,$$

where Ω is the true variance/covariance matrix. In general, however, this is unknown.

- There are two important cases when Ω is known. First when $\Omega = I_T$. The second is when:

$$y_t = \frac{1}{N_t}\sum_{i=1}^{N_t} y_{ti}, \quad u_t = \frac{1}{N_t}\sum_{i=1}^{N_t} u_{ti},\ Var(u_{ti}) = \sigma^2.$$

So we have data that are averages over some other "primary data." Maybe we have income observation by county and we average the households in each county and run our regression. Since the number of people that live in each county varies, we should use something different from I_T for our variance covariance matrix. Note:

$$Var(u_t) = \frac{1}{N_t^2}\sum_{i=1}^{N_t} Var(u_{ti}) = \frac{\sigma^2}{N_t}.$$

Then:

$$\Omega = \begin{bmatrix} \frac{1}{N_1} & & 0 \\ & \ddots & \\ 0 & & \frac{1}{N_T} \end{bmatrix}.$$

Asymptotic Properties of the GLS Estimator

- Consider the following properties:
  - (B.1') $E[u] = 0$.
  - (B.2') $E[uu'] = \sigma^2\Omega$, Ω positive definite, $u^* = Pu$ is iid with $\Omega^{-1} = p'p$.
  - (B.3') X is nonstochastic.
  - (B.4') Assume:

$$lim\frac{1}{T}X'\Omega^{-1}X = \frac{1}{T}X_*'X_* = \overline{Q},\ \text{finite, nonsingular.}$$

$$(x_{it}^*)'s \text{ are uniformly bounded so } |x_{it}^*| < c < \infty.$$

Now recall:

$$\tilde{\beta} = \beta + (X'\Omega^{-1}X)^{-1}X'\Omega^{-1}u = \beta + (X_*'X_*')^{-1}X_*'u_*.$$

Thus,

$$\sqrt{T}(\tilde{\beta}-\beta) = \underbrace{(\frac{1}{T}X_*'X_*')^{-1}}_{\to \overline{Q}^{-1}} \underbrace{\frac{1}{\sqrt{T}}X_*'u_*}_{\to^d N(0,\sigma^2\overline{Q})} .$$

The first term's distribution we assumed, the second comes from a central limit theorem applied to triangular arrays. Since the terms in the variance/covariance matrix are a function of the sample size, X_* also depends on the sample size. Thus, in assumption (B.4'), we needed the assumptions of uniform boundedness and that u_* are iid to get at the CLT. Thus we have our result:

$$\sqrt{T}(\tilde{\beta}-\beta) \to^d N(0, \overline{Q}^{-1}\sigma^2\overline{QQ}^{-1}) \equiv N(0, \sigma^2\overline{Q}^{-1}).$$

- **Proposition** Given $\tilde{\sigma}^2 = \dfrac{1}{T-K}\tilde{u}'\Omega^{-1}\tilde{u}$, with $\tilde{u} = y - X\tilde{\beta}$,

$$\tilde{\beta} \to^p \beta,$$

and,

$$\tilde{\sigma}^2 \to^p \sigma^2.$$

Also:

$$\sqrt{T}(\tilde{\beta}-\beta) \to^d N(0, \sigma^2\overline{Q}^{-1}).$$

Or,

$$\tilde{\beta} \sim N(\beta, \sigma^2\frac{\overline{Q}^{-1}}{T}) \equiv N(\beta, \sigma^2(X'\Omega^{-1}X)^{-1}),$$

if σ^2 and Ω are known. If only Ω is known:

$$\tilde{\beta} \approx N(\beta, \tilde{\sigma}^2(X'\Omega^{-1}X)^{-1}).$$

Also,

$$G \cdot F \to^d \chi^2(G),$$

where F is based on the transformed variables and G is the number of restrictions. Under the null that $\beta_i = 0$,

$$\frac{\tilde{\beta}_i}{\tilde{\sigma}_{\tilde{\beta}_i}} \to^d N(0,1).$$

And finally,

$$VC(\tilde{\beta}) = \frac{1}{T-K}\tilde{\sigma}^2(X'\Omega^{-1}X)^{-1}.$$

GLS when Ω is unknown (Feasible GLS)

- Denote our estimator:

$$\tilde{\beta}_{FGLS} = (X'\hat{\Omega}^{-1}X)^{-1}X'\hat{\Omega}^{-1}y,$$

where $\hat{\Omega}$ is an estimator for Ω.

- Can we actually estimate Ω? It has $T(T+1)/2 > T$ elements so NO! But if we assume a structure on the error terms, we might be able to express it as a function of a finite number of parameter.

- The next obvious question is if $\tilde{\beta}_{GLS}$ and $\tilde{\beta}_{FGLS}$ have the same properties (at least in the limit).

- Consider two random variables such that:

$$\eta_T \to^d N(0, \Sigma),$$

$$\xi_T - \eta_T \to^p 0.$$

  Now consider:

$$\xi_T = \underbrace{\eta_T}_{\to^d N(0,\Sigma)} + \underbrace{(\xi_T - \eta_T)}_{\to^p 0} \to^d N(0, \Sigma).$$

  So ξ_T will have the same distribution if the difference in the two random variables converges in probability to zero.

- Let:

$$\eta_T = \sqrt{T}(\tilde{\beta}_{GLS} - \beta),$$

  and,

$$\xi_T = \sqrt{T}(\tilde{\beta}_{FGLS} - \beta).$$

  Then:

$$Z = \xi_T - \eta_T = \sqrt{T}(\tilde{\beta}_{FGLS} - \tilde{\beta}_{GLS}).$$

  If $Z \to^p 0$, we're golden, and the feasible estimator will have the same asymptotic properties as the true estimator.

- We will next look for sufficient conditions where this is true. It is important to note that the true and feasible estimators need not always have the same asymptotic distribution (see notes for a convoluted example).

4 Lecture 4: September 13, 2005

4.1 More on GLS

- Recall our two estimators:

$$\tilde{\beta}_{GLS} = (X'\Omega^{-1}X)^{-1}X'\Omega^{-1}y.$$

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

- **Proposition 1** Sufficient conditions for $\tilde{\beta}_{GLS} = \tilde{\beta}_{FGLS}$:
  - (1) $plim\ \frac{1}{T}X'\hat{\Omega}^{-1}X$ is finite and non-singular.
  - (2) $plim\ \frac{1}{T}X'\hat{\Omega}^{-1}u = 0$.
- **Proposition 2** Given the assumptions of proposition 1, sufficient conditions for:

$$plim\ \sqrt{T}(\tilde{\beta}_{FGLS} - \tilde{\beta}_{GLS}) = 0,$$

  ie, the true and feasible GLS estimators to have the same asymptotic distributions, are the following:

  - (1) $plim\ \frac{1}{T}X'(\hat{\Omega}^{-1} - \Omega^{-1})X = 0$.
  - (2) $plim\ \frac{1}{\sqrt{T}}X'(\hat{\Omega}^{-1} - \Omega^{-1})u = 0$.

  Note that we can write (1) as:

$$\frac{1}{T}X'\hat{\Omega}^{-1}X - \frac{1}{T}X'\Omega^{-1}X \to^p 0.$$

  So,

$$[\frac{1}{T}X'\hat{\Omega}^{-1}X]^{-1} = [\frac{1}{T}X'\Omega^{-1}X]^{-1} + \Delta_T^1.$$

  Also the second condition implies:

$$\frac{1}{\sqrt{T}}X'\hat{\Omega}^{-1}u = \frac{1}{\sqrt{T}}X'\Omega^{-1}u + \Delta_T^2.$$

  Where both the Δ terms go to zero in probability. Proof: consider writing out the above expression:

$$\sqrt{T}(\tilde{\beta}_{FGLS} - \tilde{\beta}_{GLS}) =$$

$$
\begin{aligned}
&= (\frac{1}{T}X'\hat{\Omega}^{-1}X)^{-1}\frac{1}{\sqrt{T}}X'\hat{\Omega}^{-1}u - (\frac{1}{T}X'\Omega^{-1}X)^{-1}\frac{1}{\sqrt{T}}X'\Omega^{-1}u \\
&= [(\frac{1}{T}X'\Omega^{-1}X)^{-1} + \Delta_T^1][\frac{1}{\sqrt{T}}X'\Omega^{-1}u + \Delta_T^2] - (\frac{1}{T}X'\Omega^{-1}X)^{-1}\frac{1}{\sqrt{T}}X'\Omega^{-1}u \\
&= \Delta_T^1 \underbrace{\frac{1}{\sqrt{T}}X'\Omega^{-1}u}_{\to^d RV} + \underbrace{(\frac{1}{T}X'\Omega^{-1}X)^{-1}}_{\to^d RV}\Delta_T^2 + \Delta_T^1\Delta_T^2 \to^p 0
\end{aligned}
$$

Noting again the Δ terms both go in probability to zero. QED.

- **Corrolary** It follows immediately that:

$$
\sqrt{T}(\tilde{\beta}_{FGLS} - \beta) \to^d N(0, \sigma^2 lim_{T\to\infty}(\frac{1}{T}X'\Omega^{-1}X)^{-1}).
$$

And,

$$
\tilde{\beta}_{FGLS} \approx N(\beta, \hat{\sigma}^2(X'\hat{\Omega}^{-1}X)^{-1}).
$$

With:

$$
\hat{\sigma}^2 = \frac{\hat{u}'\hat{\Omega}^{-1}\hat{u}}{T-K}, \quad \hat{u} = y - X\tilde{\beta}_{FGLS}.
$$

Or,

$$
\hat{\sigma}^2 = \frac{\hat{u}_*'\hat{u}_*}{T-K}, \quad \hat{u}_* = p\hat{u} = y_* - X_*\tilde{\beta}_{FGLS}.
$$

- **Proposition 3** Given our assumptions in proposition 1 and proposition 2, the feasible estimator for σ^2 is

$$
\hat{\sigma}_F^2 = \frac{1}{T-K}(y - X\tilde{\beta}_{FGLS})'\hat{\Omega}^{-1}(y - X\tilde{\beta}_{FGLS}).
$$

And the true estimator:

$$
\hat{\sigma}^2 = \frac{1}{T-K}(y - X\tilde{\beta}_{GLS})'\Omega^{-1}(y - X\tilde{\beta}_{GLS}).
$$

If we add the assumption:

$$
\frac{1}{T}u'(\hat{\Omega}^{-1} - \Omega^{-1})u \to^p 0,
$$

we have:

$$
\hat{\sigma}_F^2 \to^p \sigma^2,
$$

and:

$$
\hat{\sigma}^2 \to^p \sigma^2.
$$

Proof for the true estimator:

$$
\begin{aligned}
\hat{\sigma}^2 &= \frac{1}{T-K}\hat{u}_*'\hat{u}_* \\
&= \frac{1}{T-K}u_*'M_{X_*}u_* \\
& \quad [NB: \quad M_{X_*} = I - X_*(X_*'X_*)^{-1}X_*' \\
&= \frac{1}{T-K}u_*'u_* - \frac{1}{T-K}u_*'X_*(X_*'X_*)^{-1}X_*'u_* \\
&= \frac{1}{T-K}u_*'u_* - \frac{T}{T-K}\frac{1}{T}u_*'X_*(\frac{1}{T}X_*'X_*)^{-1}\frac{1}{T}X_*'u_*
\end{aligned}
$$

So consider a couple terms:

$$\frac{1}{T}X_*'u_* = \frac{1}{T}X'\Omega^{-1}u,$$

which has expected value of zero and a variance of:

$$Var(\frac{1}{T}X'\Omega^{-1}u) = \frac{1}{T}\frac{1}{T}X'\Omega^{-1}\Omega\Omega^{-1}X = \underbrace{\frac{1}{T}}_{\to 0}\underbrace{\frac{1}{T}X'\Omega^{-1}X}_{\to Q} \to 0.$$

So by Chebychev, since the expected value is zero and the variance goes to zero, the whole thing goes in probability to zero. Also,

$$\frac{1}{T-K}u_*'u_* = \underbrace{\frac{T}{T-K}}_{\to 1}\underbrace{\frac{1}{T}\sum_{t=1}^{T}u_{t*}^2}_{\to^d \sigma^2} \to^p \sigma^2,$$

by Khinchine. So back to our expression:

$$\hat{\sigma}^2 = \underbrace{\frac{1}{T-K}u_*'u_*}_{\to \sigma^2} - \underbrace{\frac{T}{T-K}}_{\to 1}\underbrace{\frac{1}{T}u_*'X_*}_{\to 0}\underbrace{(\frac{1}{T}X_*'X_*)^{-1}}_{\to Q^{-1}}\underbrace{\frac{1}{T}X_*'u_*}_{\to 0} \to \sigma^2.$$

QED.

Proof for the feasible estimator:

$$\begin{aligned}
\hat{\sigma}_F^2 &= \frac{1}{T-K}\hat{u}'\hat{\Omega}^{-1}\hat{u} \\
&= \frac{1}{T-K}u'\hat{\Omega}^{-1}M_\Omega u \\
& [NB: \quad M_\Omega = I - X(X'\hat{\Omega}^{-1}X)^{-1}X'\hat{\Omega}^{-1} \\
&= \frac{1}{T-K}u'\hat{\Omega}^{-1}u - \frac{1}{T-K}u'\hat{\Omega}^{-1}X(X'\hat{\Omega}^{-1}X)^{-1}X'\hat{\Omega}^{-1}u \\
&= \frac{1}{T-K}u'\hat{\Omega}^{-1}u - \frac{T}{T-K}\frac{1}{T}u'\hat{\Omega}^{-1}X(\frac{1}{T}X'\hat{\Omega}^{-1}X)^{-1}\frac{1}{T}X'\hat{\Omega}^{-1}u
\end{aligned}$$

Again, consider the term:

$$\frac{1}{T}u'\hat{\Omega}^{-1}X \to^p O_p(1),$$

by assumption (?) Also, write the first term:

$$\frac{1}{T-K}u'\hat{\Omega}^{-1}u = \underbrace{\frac{T}{T-K}(\frac{1}{T}u'\hat{\Omega}^{-1}u - \frac{1}{T}u'\Omega^{-1}u)}_{\to^p\ 0} + \frac{T}{T-K}\underbrace{\frac{1}{T}u'\Omega^{-1}u}_{\to^p\ \sigma^2}.$$

So back to our expression:

$$\hat{\sigma}_F^2 = \underbrace{\frac{1}{T-K}u'\hat{\Omega}^{-1}u}_{\to^p \sigma^2} - \underbrace{\frac{T}{T-K}}_{\to 1}\underbrace{\frac{1}{T}u'\hat{\Omega}^{-1}X}_{\to^p\ O_p(1)}\underbrace{(\frac{1}{T}X'\hat{\Omega}^{-1}X)^{-1}}_{\to Q^{-1}}\underbrace{\frac{1}{T}X'\hat{\Omega}^{-1}u}_{\to^p\ 0} \to \sigma^2.$$

QED.

5 Lecture 5: September 15, 2005

5.1 Autocorrelation and Heteroskedasticity

- First some preliminaries. Consider a random variable, z_t. Then using the lag operator, we have:
$$Lz_t = z_{t-1}, \quad L^2 z_t = z_{t-2}.$$
- Define the following two lag polynomials:
$$A(L) = 1 - a_1 L - \cdots - a_p L^p,$$
$$B(L) = 1 + b_1 L + \cdots + b_q L^q.$$
- Then consider the following structure for the error terms of a model:
$$u_t - a_1 u_{t-1} - \cdots - a_p u_{t-p} = \epsilon_t + b_1 \epsilon_{t-1} + \cdots + b_q \epsilon_{t-q},$$
or,
$$A(L) u_t = B(L) \epsilon_t.$$
This is an $ARMA(p,q)$ process.

The $AR(1)$ Process

- Consider an $AR(1)$ or $ARMA(1,0)$ process:
$$u_t = \rho u_{t-1} + \epsilon_t, \quad E[\epsilon_t] = 0, \; E[\epsilon_t^2] = \sigma_\epsilon^2, \; E[\epsilon_t \epsilon_s] = 0 \; t \neq s.$$
Iterative substitution yields:
$$u_t = \rho^s u_{t-s} + \sum_{i=0}^{s-1} \rho^i \epsilon_{t-i}.$$
If we assume $|\rho| < 1$, the process is stationary and the first term goes to zero as s goes to infinity. So:
$$u_t = \sum_{i=0}^{\infty} \rho^i \epsilon_{t-i}.$$
Another way we could get at this is to write out the homogenous and particular solutions and again note the homogenous solution goes to zero.
- Thus,
$$E[u_t] = 0,$$
and the variance:
$$E[u_t^2] = E[(\epsilon_t + \rho \epsilon_{t-1} + \rho^2 \epsilon_{t-2} + \ldots)(\epsilon_t + \rho \epsilon_{t-1} + \rho^2 \epsilon_{t-2} + \ldots)].$$

but since $E[\epsilon_t^2] = \sigma_\epsilon^2$ and $E[\epsilon_t\epsilon_s] = 0$ for all $t \neq s$, we have:

$$E[u_t^2] = \sigma^2 + \rho^2\sigma^2 + \rho^4\sigma^2 + \cdots = \sigma_\epsilon^2 \sum_{i=0}^{\infty} \rho^{2i} = \sigma_\epsilon^2 \frac{1}{1-\rho^2}.$$

- Finally, the covariance:

$$\begin{aligned}
E[u_t u_{t+s}] &= E\left[(\sum_{i=1}^{\infty} \rho^i \epsilon_{t-i})(\sum_{j=1}^{\infty} \rho^j \epsilon_{t+s-j}) \right]. \\
&= \sum_{i=0}^{\infty}\sum_{j=0}^{\infty} \rho^{i+j} \underbrace{E[\epsilon_{t-i}\epsilon_{t+s-j}]}_{=0 \ except \ i=j-s}. \\
&= \sum_{i=0}^{\infty} \rho^{i+(i+s)} E[\epsilon_{t-i}\epsilon_{t+s-(i+s)}]. \\
&= \sum_{i=0}^{\infty} \rho^{2i+s} \underbrace{E[\epsilon_{t-i}\epsilon_{t-i}]}_{\sigma_\epsilon^2}. \\
&= \sigma_\epsilon^2 \rho^s \sum_{i=0}^{\infty} \rho^{2i}. \\
&= \sigma_\epsilon^2 \frac{\rho^{|s|}}{1-\rho^2}.
\end{aligned}$$

- Note the variance is constant and the covariance has limited memory (goes to zero for a larger and larger time inverval). This process is thus ergodic because it is stationary and has limited memory. Since the covariance is still not constant across time, the process is also called "weakly stationary."

- So consider the model:

$$y = X\beta + u, \;\; u_t = \rho u_{t-1} + \epsilon_t, \; E[u] = 0.$$

Thus,

$$E[uu'] = \sigma_u^2\Omega = \sigma_u^2 \begin{bmatrix} 1 & \rho & \rho^2 & \ldots & \rho^{T-1} \\ \rho & 1 & \rho & \ldots & \rho^{T-2} \\ \rho^2 & & \ddots & & \vdots \\ \vdots & & & \ddots & \vdots \\ \rho^{T-1} & \rho^{T-2} & \ldots & \ldots & 1 \end{bmatrix}.$$

- For convenience write this as $E[uu'] = \sigma_u^2\Omega = \sigma_\epsilon^2 W$ with $W = \frac{\sigma_u^2}{\sigma_\epsilon^2}\Omega$.

- Recall our estimator:

$$\hat{\beta}_{GLS} = (X'\Omega^{-1}X)^{-1}X'\Omega^{-1}y = (X'W^{-1}X)^{-1}X'W^{-1}y,$$

  which is found by:

$$Min_{\beta}\ \{(y - X\beta)'W^{-1}(y - X\beta)\}.$$

  And if $p'p = W^{-1}$, we have:

$$Min_{\beta}\ \{(Py - PX\beta)'(Py - PX\beta)\},$$

  which is the OLS estimator using the transformed model.

- We could also find the estimator using maximum likelihood:

$$Min_{\beta}\ \{|\Omega|^{1/T}(y - X\beta)'\Omega^{-1}(y - X\beta)\}.$$

  See Judge, Pg 284. So the MLE equals the GLS estimator.

- Or we could have run regular OLS on the original data:

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

  Note $E[\hat{\beta}] = 0$, so the OLS estimator is still unbiased since $E[u|X] = 0$. However,

$$VC(\hat{\beta}) = E[(X'X)^{-1}X'uu'X(X'X)^{-1}] = \sigma_u^2(X'X)^{-1}X'\Omega X(X'X)^{-1}.$$

  Thus, the only effect this would have is for hypothesis testing. The variance of our estimators would be different.

- So consider the model again:

$$y_t = x_{t\cdot}\beta + u_t.$$

  Consider lagging once and mulitplying by ρ:

$$\rho y_{t-1} = \rho x_{t-1\cdot} + \rho u_{t-1}.$$

  Subtracting we have:

$$y_t - \rho y_{t-1} = (x_{t\cdot} - \rho x_{t-1\cdot})\beta + \underbrace{u_t - \rho u_{t-1}}_{\epsilon_t},\ t = 2 \ldots T.$$

  So we can tranform the data (for observations two onward) like this and get nicely behaved errors. What about the first observation? Consider:

$$\sqrt{1-\rho^2}y_1 = \sqrt{1-\rho^2}x_{1\cdot}\beta + \sqrt{1-\rho^2}u_1.$$

  Then,

$$Var(\sqrt{1-\rho^2}u_1) = (1-\rho^2)\frac{\sigma_\epsilon^2}{1-\rho^2} = \sigma_\epsilon^2.$$

So if we can find a transformation that does this to our data, we're golden. The errors will have nice properties and we can run OLS on the transformed data.

- So consider the following W:

$$W = \frac{1}{1-\rho^2}\begin{bmatrix} 1 & \rho & \rho^2 & \dots & \rho^{T-1} \\ \rho & 1 & \rho & \dots & \rho^{T-2} \\ \rho^2 & & \ddots & & \vdots \\ \vdots & & & \ddots & \vdots \\ \rho^{T-1} & \rho^{T-2} & \dots & \dots & 1 \end{bmatrix}.$$

This implies:

$$W^{-1} = \begin{bmatrix} 1 & -\rho & 0 & \dots & 0 \\ -\rho & 1+\rho^2 & -\rho & & \vdots \\ 0 & -\rho & \ddots & \ddots & 0 \\ \vdots & & \ddots & 1+\rho^2 & -\rho \\ 0 & \dots & 0 & -\rho & 1 \end{bmatrix}.$$

Check that $WW^{-1} = I$. We can also see that:

$$p = \begin{bmatrix} \sqrt{1-\rho^2} & 0 & & \dots & 0 \\ -\rho & 1 & 0 & & \vdots \\ 0 & -\rho & 1 & & \\ \vdots & \ddots & \ddots & \ddots & 0 \\ 0 & \dots & 0 & -\rho & 1 \end{bmatrix}.$$

Check that $p'p = W^{-1}$. And finally, as an example, consider:

$$py = \begin{bmatrix} \sqrt{1-\rho^2} & 0 & & \dots & 0 \\ -\rho & 1 & 0 & & \vdots \\ 0 & -\rho & 1 & & \\ \vdots & \ddots & \ddots & \ddots & 0 \\ 0 & \dots & 0 & -\rho & 1 \end{bmatrix} \begin{bmatrix} y_1 \\ \vdots \\ y_T \end{bmatrix} = \begin{bmatrix} \sqrt{1-\rho^2}y_1 \\ y_2 - \rho y_1 \\ y_3 - \rho y_2 \\ \vdots \\ y_T - \rho y_{T-1} \end{bmatrix}.$$

So it works!

- So you can apply OLS to the transformed model and you get the GLS estimator:

$$\hat{\beta}_{GLS} = (X'\Omega^{-1}(\rho)X)^{-1}X'\Omega^{-1}(\rho)y,$$

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

- Next we determine methods of estimating ρ.

6 Lecture 6: September 20, 2005

6.1 More on AR(1)

- Recall our model:

$$y_t = X_{t\cdot}\beta + u_t, \;\; u_t = \rho u_{t-1} + \epsilon_t, \;\; t = 1\ldots, T.$$

Which we transform into:

$$y_t - \rho y_{t-1} = (X_{t\cdot} - \rho X_{t-1\cdot})\beta + \epsilon_t, \;\; t = 2\ldots T.$$

- So we need to estimate ρ. We will now consider several methods.

1. Cochrane Orcutt Method

- Run OLS of y on X and estimate β, which will be consistent and unbiased (though inefficient). Calculate the residuals, $\hat{u}$. Then consider:

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

Plug $\hat{\rho}$ into the transformed model and run OLS to get an estimate for β.

- Note in the first step, we are solving:

$$Min_\rho \; \{\sum_{t=2}^{T}[\underbrace{(y_t - X_{t\cdot}\hat{\beta})}_{u_t} - \rho\, \underbrace{(y_{t-1} - X_{t-1\cdot}\hat{\beta})}_{u_{t-1}}]^2\}.$$

And then in the second step, we are solving:

$$Min_\beta \; \{\sum_{t=2}^{T}[(y_t - \hat{\rho} y_{t-1}) - (X_{t\cdot} - \hat{\rho} X_{t-1\cdot})\beta]^2\}.$$

And we could ITERATE between the two, increasing our efficiency as we go (hopefully). This is called *Iterated Cochrane Orcutt.*

- But this is completely EQUIVALENT to :

$$Min_{\beta,\rho} \; \{\sum_{t=2}^{T}[(y_t - \rho y_{t-1}) - (X_{t\cdot} - \rho X_{t-1\cdot})\beta]^2\},$$

which is Non-Linear Least Squares.

2. Hildreth-Lu Method

- We know that $\rho \in (-1, 1)$. Thus, pick a grid of ρ's and minimize the objective with respect to β for each ρ. This is an inelegant way to minimize the objective but it works.

3. Durbin Method

- Consider rewriting the model:

$$y_t = \rho y_{t-1} + X_{t\cdot}\beta - X_{t-1\cdot}\underbrace{\rho\beta}_{\gamma} + \epsilon_t,$$

so regress y on the lag of y, X, and the lag of X. The coefficient of the lag of y is your estimate of ρ.

4. Maximum Likelihood Method

- Recall our objective function:

$$Min\ \{\sum_{t=2}^{T}[\underbrace{y_t - \rho y_{t-1}}_{y_t^*} - (\underbrace{X_{t\cdot} - \rho X_{t-1\cdot}}_{X_t^*})\beta]^2\}.$$

The likelihood function is thus:

$$L = const + \frac{T}{2}ln(\frac{1}{\sigma_\epsilon^2}) + \underbrace{\frac{1}{2}ln(1-\rho^2)}_{\phi} - \frac{1}{2\sigma_\epsilon^2}(y^* - X^*\beta)'(y^* - X^*\beta).$$

Where the ϕ term is the only difference between this and the nonlinear least squares ML estimator and it comes from the determinant of the Jacobian of the transformation.

- Example for an AR(4) model:

$$y_t = X_{t\cdot}\beta + u_t, \;\; u_t = \rho u_{t-4} + \epsilon_t.$$

Transformed model

$$y_t - \rho y_{t-4} = (X_{t\cdot} - \rho X_{t-4\cdot})\beta + \epsilon_t, \;\; t = 5 \ldots T.$$

And for the first 4 observations:

$$\sqrt{1-\rho^2}y_t = \sqrt{1-\rho^2}X_{t\cdot}\beta + \sqrt{1-\rho^2}u_t, \;\; t = 1 \ldots 4.$$

We can run the analogous tests on this model as above.

6.2 Tests for Autocorrelation

Durbin Watson Test (DW)

- Consider the model:

$$y_t = X_{t\cdot}\beta + u_t, \quad u_t = \rho u_{t-1} + \epsilon_t, \quad X_{t,1} = 1, (X_{t,i})'s \text{ are nonstochastic.}$$

- Run OLS on the model and save the residuals, $\hat{u}$. Calculate:

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

Multiplying out, we have:

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

- Now consider three cases:
  - Case (1): $\rho = 0$. Then, the last term is like a covariance between successive u_t's which is zero.
  $$d \approx 2.$$
  - Case (2): $\rho \approx 1$. Then $u_t = u_{t-1}$ and
  $$d \approx 0.$$
  - Case (3): $\rho \approx -1$. Then:
  $$d \approx 4.$$

- So, in general, $d \in (0, 4)$. Under the null, $H_0 : \rho = 0$, this means $d = 2$. So we would fail to reject the null for a realized d of around 2, but if it was less than d_α or greater than $4 - d_\alpha$, we would reject. The rejection region, however, depends on the sample X's. So some statistical packages will calculate the bounds on the fly. Textbooks give ranges for the critical values so if your realized d is below the smallest critical value or bigger than the largest, you can reject. If it's in between, the test is indeterminant. Usually, the software will give you just one set of critical values based on your data so this indeterminancy goes away.

Adjusted Durbin Watson Test (h statistic)

- If there is a lagged dependent variable in your regression, the DW statistic above is invalid. Consider the model:

$$y_t = y_{t-1}\beta_1 + X_{t\cdot}\beta_1^* + u_t.$$

Calculate d as in the DW test and then calculate:

$$h = (1 - \frac{1}{2}d)\sqrt{\frac{T}{1 - T * \hat{Var}(\hat{\beta}_1)}} \approx N(0,1),$$

under the null of no serial correlation. Note the variance term is the estimated variance of the OLS estimator for β_1, the coefficient on your lagged y.

7 Lecture 7: September 22, 2005

7.1 More Tests for Autocorrelation

Breusch-Godfrey Test (LM Test)

- In this test, we only have to estimate the model under the null hypothesis (no autocorrelation). Consider the model:

$$y_t = X_{t\cdot}\beta + u_t,$$

where the errors are either:

$$u_t = a_1 u_{t-1} + \cdots + a_p u_{t-p} + \epsilon_t, \quad [AR(p)]$$

or,

$$u_t = \epsilon_t + b_1 \epsilon_{t-1} + \cdots + b_p \epsilon_{t-p}, \quad [MA(p)].$$

The null is: $H_0 : u_t = \epsilon_t$. So we calculate:

$$LM = \left(\frac{\hat{u}' X_0 (X_0' X_0)^{-1} X_0' \hat{u}}{\hat{u}'\hat{u}/T} \right) \sim \chi^2(p).$$

Where $\hat{u}$ are the OLS residuals, and:

$$X_0 = [X_{t1} \ldots X_{tK}, \hat{u}_{t-1}, \ldots, \hat{u}_{t-p}].$$

- To give an intuition for this test statistic, consider the model $y = X\beta + u$ and OLS fitted values:

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

Then the regression sum of squares is:

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

Thus the numerator above is the RSS from a regression of the residuals on the X's and the lagged u's. By definition the X's and the residuals should be orthogonal, so the top gives us the additional explanatory power of the p lags of the residuals. If this is large, we should reject the null.

Box Pierce Test

- Consider the model:

$$y_t = X_{t\cdot}\beta + u_t, \quad u_t = \rho u_{t-1} + \epsilon_t.$$

Then:

$$\hat{u}_t = y_t - X_{t\cdot}\hat{\beta}.$$

Calculate:

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

or more generally, the higher order correlations are:

$$\hat{\rho}_j = \frac{\sum_{t=j+1}^{T} \hat{u}_t \hat{u}_{t-j}}{\sum_{t=1}^{T} \hat{u}_t^2}.$$

Then consider:

$$Q = T \sum_{i=1}^{p} \hat{\rho}_j^2 \sim \chi^2(p),$$

where p is subjectively choosen as the number of lags that you think the error terms are serially correlated.

7.2 Heterskedasticity

- Consider the model:

$$y_t = \beta_1 + \beta_2 x_{t2} + \cdots + \beta_K x_{tK} + u_t, \quad E[u_t] = 0,$$

and,

$$E[u_t^2] = h(X, Z, \alpha).$$

So the second moment of the errors depends possibly on the data, some parameters which may or may not already be in the model, and possibly external variables, Z.

- OLS will be unbiased and consistent in this case, but the estimators will be inefficient and statistical tests will be wrong because of the variance structure is no longer $\sigma^2 I_T$.

- So consider a general setup where we only have heteroskedasticity (no autocorrelation) and we write the model:

$$y_t = X_{t\cdot}\beta + u_t, \; E[u_t] = 0, \; E[u_t^2] = \sigma^2 \omega_t^2.$$

Which we can write in matrix form as:

$$y = X\beta + u, \; E[u] = 0, \; E[uu'] = \sigma^2 \Omega = diag(\omega_t^2).$$

We could estimate this by maximum likelihood as follows. Consider the log likelihood function:

$$ln(L) = const + \frac{1}{2} ln(det(\sigma^2 \Omega)^{-1}) - \frac{1}{2}(y - X\beta)' \Omega^{-1} (y - X\beta),$$

with:

$$det(\sigma^2\Omega)^{-1} = \begin{bmatrix} \frac{1}{\sigma^2\omega_1^2} & & 0 \\ & \ddots & \\ 0 & & \frac{1}{\sigma^2\omega_T^2} \end{bmatrix}.$$

If $\omega_t = \omega_t(\alpha, \beta)$, we would not get the same estimator as under GMM of NLLS.

- If we know ω_t, we could calculate:

$$\hat{\beta}_{GLS} = (X'\Omega^{-1}X)^{-1}X'\Omega^{-1}y = \left[\sum_{t=1}^{T} \omega_t^{-2} X'_{t\cdot} X_{t\cdot}\right]^{-1} \sum_{t=1}^{T} \omega_t^{-2} X'_{t\cdot} y_t.$$

  Or using the transformed model:

$$\frac{y_t}{\omega_t} = \frac{x_{t1}}{\omega_t}\beta_1 + \cdots + \frac{x_{tK}}{\omega_t}\beta_K + \frac{u_t}{\omega_t}.$$

- So if we know Ω, we're golden. What if Ω is unknown? Suppose we know that $\omega_t = \omega_t(\alpha, \beta)$, simply a function of parameters of the model. Well, we could estimate those parameters by OLS and find $\hat{\omega}$ and then compute our Feasible GLS estimator. But even here we are assuming we know the functional form of ω.

- We could also do NonLinear Least Squares. Consider the model:

$$\frac{y_t}{\omega_t(\alpha,\beta)} = \frac{X_{t\cdot}}{\omega_t(\alpha,\beta)}\beta + \frac{u_t}{\omega_t(\alpha,\beta)}.$$

  Then,

$$q_t(\alpha, \beta, y_t, x, z_t) = u_t^* = \frac{y_t}{\omega_t(\alpha,\beta)} - \frac{X_{t\cdot}}{\omega_t(\alpha,\beta)}\beta.$$

  And calculate:

$$Q(\theta) = \sum_{t=1}^{T} q_t(\alpha, \beta, y_t, x_t, z_t)^2.$$

  The NLLS estimator solves:

$$\frac{\partial Q(\hat{\theta})}{\partial \theta} = 0.$$

- However, even if we don't know the functional form of ω, we can still estimate the variance of the coefficients we get from running OLS when there is heterskedasticity. Next time.

8 Lecture 8: September 27, 2005

8.1 More on Heteroskedasticity

Estimating Parameter Efficiency

- Consider the model:

$$y = X_{t\cdot}\beta + u_t,\ E[u_t] = 0,\ E[u_t^2] = \sigma_t^2,$$

or,

$$y = X\beta + u,\ E[u] = 0\ E[uu'] = \sigma^2\Omega = \Sigma.$$

We could estimate with OLS yielding:

$$\hat{\beta}_{OLS} = (X'X)^{-1}X'y,\ \ VC(\hat{\beta}) = (X'X)^{-1}X'\Sigma X(X'X)^{-1}.$$

- The problem is now, can we estimate $X'\Sigma X$ without knowing all the σ's ?
- Consider:

$$\psi = \frac{1}{T}X'\Sigma X = \frac{1}{T}\sum_{t=1}^{T}\sigma_t^2 X_{t\cdot}'X_{t\cdot}.$$

Or,

$$\psi = E[\underbrace{\frac{1}{T}\sum_{t=1}^{T}u_t^2 X_{t\cdot}'X_{t\cdot}}_{\hat{\psi}}].$$

Then, $E[\hat{\psi}] = \psi$.

- For example, consider the case of one regressor:

$$\hat{\psi} = \frac{1}{T}\sum_{t=1}^{T}u_t^2 x_t^2.$$

And,

$$E[\hat{\psi}] = \psi = \frac{1}{T}\sum_{t=1}^{T}\sigma_t^2 x_t^2.$$

- So White (1980) proposed the following method for estimating the standard errors of the coefficients without knowing the actual variance structure of the errors: Denote:

$$\tilde{\psi} = \frac{1}{T}\sum_{t=1}^{T}\tilde{u}_t^2 X_{t\cdot}'X_{t\cdot},\ \ \tilde{u}_t = y_t - X_{t\cdot}\hat{\beta}_{OLS}.$$

It can be shown:

$$\tilde{\psi}_T - \psi \to^p 0,$$

consistent. These are called the Robust Standard Errors.

- If you want find the standard errors when you suspect the errors suffer from both heteoskedasticity and autocorrelation, then calculate a HAC estimator, (Heteroskedastic Autocorrelation Consistent Estimator). See notes. There other methods as well including Newey West (1987), Andrews (1981), and Potsher/Prucha (1997).

8.2 Estimating Under Alternative Specifications

- In this section we consider various forms of heteroskedasticity and consider ways in which we could estimate the model. The model is always:

$$y = X_{t\cdot}\beta + u_t,\ E[u_t] = 0,\ E[u_t^2] = \sigma_t^2.$$

Specification A

- Suppose $E[u_t] = 0$ and $E[u_t^2] = \sigma_t^2 = \sigma^2 h(z_t)$, where $h(\cdot)$ is a known function.

- Here, since we know the form of σ_t^2, at least up to a constant, we can just divide the data, (y, X), and errors, u_t, by the square root of $h(z_t)$ and run OLS on the transformed model. Since we actually know the true variance structure, this is really true GLS. An example might be the variance of the errors growing with the income level of a household.

Specification B

- Suppose $E[u_t] = 0$ and $E[u_t^2] = \sigma_t^2 = \sigma^2 (z_{t\cdot}\alpha)^p$, for $p = 1$ or 2. See Judge, 1985. So if $p = 1$,

$$E[u_t^2] = \sigma^2 z_{t\cdot}\alpha = z_{t\cdot}\alpha^*,\quad \alpha^* = \sigma^2\alpha.$$

We could then write:

$$u_t^2 = z_{t\cdot}\alpha^* + \epsilon_t,\ E[\epsilon_t] = 0.$$

Running OLS on this model will give us a consistent estimate of α^*. In particular:

$$\hat{\alpha}^* = [\sum_t z_{t\cdot}' z_{t\cdot}]^{-1}[\sum_t z_{t\cdot}' u_t^2].$$

Then we can estimate the true GLS estimator with:

$$\hat{\beta}_{GLS} = [\sum_t (z_{t\cdot}\alpha^*)^{-1} x_{t\cdot}' x_{t\cdot}]^{-1}[\sum_t (z_{t\cdot}\alpha^*)^{-1} x_{t\cdot}' y_t].$$

And the feasible GLS estimator:

$$\hat{\beta}_{FGLS} = [\sum_t (z_{t\cdot}\hat{\alpha}^*)^{-1} x_{t\cdot}' x_{t\cdot}]^{-1}[\sum_t (z_{t\cdot}\hat{\alpha}^*)^{-1} x_{t\cdot}' y_t],$$

with,

$$\hat{\alpha}^* = [\sum_t z'_{t\cdot} z_{t\cdot}]^{-1}[\sum_t z'_{t\cdot}\hat{u}_t^2],$$

and the $\hat{u}$'s are the OLS residuals.

- So we have several steps: First regress y on X using OLS to obtain residuals. Square the residuals and regress these on the z's. Obtain a consistent and unbiased estimator of α^*. Transform the data by $z_t\hat{\alpha}^*$. Run OLS again on the transformed data to get a consistent, unbiased, and efficient estimate of β, our FGLS estimator.

Specification C

Suppose $E[u_t] = 0$ and $E[u_t^2] = \sigma_t^2 = \sigma^2(X_{t\cdot}\beta)^p$. Since $\hat{\beta}_{OLS}$ is unbiased and consistent even with heteroskedastic errors, we can just estimate with OLS. Then,

$$E[u_t^2] = \sigma^2(X_{t\cdot}\hat{\beta}_{OLS})^p.$$

Scale through by the square root of this and run OLS on the transformed model and we're golden. We could also do NLLS:

$$Min_\beta \ \sum_t \left(\frac{y_t - X_t\beta}{\sqrt{(X_t\beta)^p}}\right)^2.$$

Specification D

- Suppose now the data is in blocks where the variance is constant, but overall there is heteroskedasticity. Suppose there are $i = 1 \ldots m$ subgroups with,

$$E[u_i] = 0, \ \ E[u_i u_i'] = \sigma_i^2 I_{T_i}, \ \ E[u_i u_j'] = 0 \ \forall \ i \neq j.$$

Then $E[u] = 0$ and $E[uu'] = diag(\sigma_1^2, \ldots, \sigma_m^2)$ with zeros everywhere else. Our usual GLS estimator is:

$$\hat{\beta}_{GLS} = [\sum_{i=1}^m \sigma_i^{-2} x'_{i\cdot} x_{i\cdot}]^{-1}[\sum_{i=1}^m \sigma_i^{-2} x'_{i\cdot} y_i].$$

But how do we find the σ_i's ? Just run OLS on each subgroup. Calculate $\hat{u}_i$ for $i = 1 \ldots m$. Then:

$$\hat{\sigma}_i^2 = \frac{1}{T_i - K}\hat{u}_i'\hat{u}_i.$$

This is a consistent estimate of the σ_i^2's. Divide all group i data by $\hat{\sigma}_i$ and then run OLS on the transformed model.

8.3 Tests for Heteroskedasticity

- Consider the null hypothesis: $H_0 : \sigma_t^2 = \sigma^2$, ie homoskedastic errors.

Test A: White Test

- Run the model with OLS and save the residuals. Run OLS on:

  $$\hat{u}_t^2 = a + f(x_{t1}, x_{t2}, \ldots, x_{tp}) + \eta_t.$$

  where the function f may include both the individual x's and cross products. Then our test statistic is:

  $$T * R^2 \sim \chi^2(p-1).$$

  We reject the null for a large value because this means the x's are still explaining some of the variation in the residuals.

9 Lecture 9: September 29, 2005

9.1 More Tests for Heteroskedasticity

Test B: Breusch Pagan Test

- Also attribured to Godrey. This a LM test.
- We assume $E[u_t] = 0$ and $E[u_t^2] = \sigma_t^2 = h(z_{t\cdot}\alpha)$, where $z_{t\cdot}$ is $1xs$ and contains an intercept, possibly the x's and possibly other exogenous variables. α is $sx1$.
- Assume $u_t \sim$ normal and our null is:

$$H_0 : \alpha_2 = \alpha_3 = \cdots = \alpha_s = 0,$$

or the errors are homoskedastic.

- The test statistic is thus:

$$LM = \frac{RSS}{2\hat{\sigma}^4} \sim \chi^2(s-1), \;\; \hat{\sigma}^2 = \frac{1}{T}\sum_{t=1}^{T} \hat{u}_t^2,$$

where $\hat{u}$ are the OLS residuals and the Regression Sum of Squares is from the regression:

$$\hat{u}_t^2 = z_{t\cdot}\alpha + v_t.$$

- We can weaken the normality assumption to just iid if we impose Krenker (1981).
- Note we DO NOT need a functional form of $h(\cdot)$ because we estimate the model under the null which means $E[u_t^2] = h(\alpha_1) \equiv$ constant.

Test C: Goldfeld Quandt Test

- Consider splitting our sample, $(1 \ldots T)$, into two subsamples of equal size, $\frac{T-r}{2}$, where r is choosen arbitrarily as the set of middle observations. Suppose we posit that the variance is increaseing in x_5. Then order the data by x_5, pull off the first $(T-r)/2$ and the last $(T-r)/2$ observations and calculate the Error Sum of Squares for each subsample. Under the alternative hypothesis of heteoskedasticity, (in general):

$$\sigma_1^2 \leq \sigma_2^2 \leq \cdots \leq \sigma_T^2.$$

Under the null of homoskedasticity, $\sigma_1^2 = \sigma_2^2$,

$$\frac{ESS_i}{\sigma^2} = \frac{u_i'[I - X_i(X_i'X_i)^{-1}X_i']u_i}{\sigma^2} = \frac{\hat{u}_i'\hat{u}_i}{\sigma^2} \sim \chi^2(\frac{T-r}{2} - K).$$

Thus our test statistic is:

$$GQ = \frac{ESS_2}{ESS_1} \sim F(\frac{T-r}{2} - K, \frac{T-r}{2} - K),$$

because we have a ratio of independent χ^2's.

- For large values of GQ, reject the null, but this is only due to our ordering of the observations from small (expected) variance to large. Under the null, $GQ \approx 1$.

9.2 Seemingly Unrelated Regression

- This analysis is due to Zellner (1492). Suppose we have two equations:

$$y_{t1} = f_1(x_{t1}, \beta_1) + u_{t1},$$

$$y_{t2} = f_2(x_{t2}, \beta_2) + u_{t2}.$$

  Or in matrix form:

$$y_{\cdot 1} = X_1\beta_1 + u_{\cdot 1}, \;\; y_{\cdot 2} = X_2\beta_2 + u_{\cdot 2}.$$

- So $u_{\cdot 1}$ is the vector of all disturbances of equation 1. Then:

$$E[u_{\cdot 1}] = E[u_{\cdot 2}] = 0,$$

$$E[u_{\cdot 1}u'_{\cdot 1}] = \sigma_{11}I_T, \quad E[u_{\cdot 2}u'_{\cdot 2}] = \sigma_{22}I_T.$$

- We could estimate all the equations separately by OLS (if linear) and all the estimators would be unbiased, consistent, and "efficient." However, they would only be efficient relative to the information set regarding the individual equations. Take the two equations together, and the information set gets larger so the most efficient way to estimate our coefficients will no longer be estimating each equation separately. Hence the expression, "seemingly unrelated."

- So suppose we also have:

$$E[u_{\cdot 1}u'_{\cdot 2}] = \sigma_{12}I_T.$$

  So we have some covariance across equations but NOT across time. Just contemporaneous correlations in the equations.

- So now lets generalize to m equations such that:

$$E[u_{\cdot i}] = 0, \;\; E[u_{\cdot i}u'_{\cdot j}] = \sigma_{ij}I_T \; \forall \; (i, j).$$

- We might get a better (more efficient estimate), if we stack the data. So consider the following stacking:

$$y = \begin{bmatrix} y_{\cdot 1} \\ \vdots \\ y_{\cdot m} \end{bmatrix}.$$

$$X = \underbrace{\begin{bmatrix} X_1 & & 0 \\ & \ddots & \\ 0 & & X_m \end{bmatrix}}_{\textit{block diagonal}}.$$

$$\beta = \begin{bmatrix} \beta_1 \\ \vdots \\ \beta_m \end{bmatrix}.$$

$$u = \begin{bmatrix} u_{\cdot 1} \\ \vdots \\ u_{\cdot m} \end{bmatrix}.$$

So our model is $y = X\beta + u$ with $E[u] = 0$. However, for the variance structure, we have;

$$E[uu'] = E[\begin{pmatrix} u_{\cdot 1} \\ \vdots \\ u_{\cdot m} \end{pmatrix} * \begin{pmatrix} u'_{\cdot 1} & \dots & u'_{\cdot m} \end{pmatrix}] = E \begin{bmatrix} u_{\cdot 1}u'_{\cdot 1} & \dots & u_{\cdot 1}u'_{\cdot m} \\ \vdots & \ddots & \vdots \\ u_{\cdot m}u'_{\cdot 1} & \dots & u_{\cdot m}u'_{\cdot m} \end{bmatrix}.$$

Which we can write:

$$E[uu'] = \begin{bmatrix} \sigma_{11}I_T & & \dots & & \sigma_{1m}I_T \\ & \ddots & & & \\ \vdots & & \sigma_{ij}I_T & & \vdots \\ & & & \ddots & \\ \sigma_{m1}I_T & & \dots & & \sigma_{mm}I_T \end{bmatrix} = \Sigma \otimes I_T = \Omega,$$

with,

$$\Sigma_{mxm} = \begin{bmatrix} \sigma_{11} & \dots & \sigma_{1m} \\ \vdots & \ddots & \vdots \\ \sigma_{m1} & \dots & \sigma_{mm} \end{bmatrix}.$$

- A reminder about kronecker products. If A is kxm and B is another matrix, then:

$$A \otimes B = \begin{bmatrix} a_{11}B & \dots & a_{1m}B \\ \vdots & \ddots & \vdots \\ a_{k1}B & \dots & a_{km}B \end{bmatrix}.$$

- So back to our variance/covariance matrix above,

$$E[uu'] = \Sigma \otimes I_T = \Omega.$$

We can think of things a bit differently if we consider the matrix:

$$U = \begin{bmatrix} u_{\cdot 1} & \dots & u_{\cdot m} \end{bmatrix} = \begin{bmatrix} u_{1\cdot} \\ \vdots \\ u_{T\cdot} \end{bmatrix}.$$

So $u_{t\cdot}$ is $1xm$ and represents the disturbances of the t^{th} observation of each of the m equations. Then,

$$E[u'_{t\cdot}u_{s\cdot}] = \begin{cases} \Sigma, & t = s \\ 0, & t \neq s \end{cases}$$

- So what's next? Well, we could clearly do GLS on the model using our variance covariance matrix, Ω. Thus we have:

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

  Or we could rewrite this:

  $$\hat{\beta}_{GLS} = (X'[\Sigma \otimes I_T]^{-1}X)^{-1}X'[\Sigma \otimes I_T]^{-1}y.$$

  $$\hat{\beta}_{GLS} = (X'[\Sigma^{-1} \otimes I_T]X)^{-1}X'[\Sigma^{-1} \otimes I_T]y.$$

  Note the second way of writing the estimator is computationally more efficient since once you take the kronecker product, you have a huge matrix to invert.

- We can write the GLS estimator explicitly as:

  $$\hat{\beta}_{GLS} = \begin{bmatrix} \sigma^{11}X_1'X_1 & \dots & \sigma^{1m}X_1'X_m \\ \vdots & \ddots & \vdots \\ \sigma^{m1}X_m'X_1 & \dots & \sigma^{mm}X_m'X_m \end{bmatrix}^{-1} * \begin{bmatrix} \sum_{i=1}^m \sigma^{1i}X_1'y_{\cdot i} \\ \vdots \\ \sum_{i=1}^m \sigma^{mi}X_m'y_{\cdot i} \end{bmatrix}.$$

- If $\Sigma = I_T$, the GLS estimator just collapses to the OLS estimator.

- Of course, usually we cannot estimate via GLS, but instead form our Feasible GLS estimator as:

  $$\hat{\beta}_{FGLS} = (X'[\hat{\Sigma}^{-1} \otimes I_T]X)^{-1}X'[\hat{\Sigma}^{-1} \otimes I_T]y.$$

  But now we need estimates of the σ_{ij}'s. Consider:

  $$\sigma_{ij} = E[u_iu_j] = \frac{1}{T}\sum_t E[u_{ti}u_{tj}],$$

  which we estimate with:

  $$\hat{\sigma}_{ij} = \frac{1}{T}\sum_t \hat{u}_{ti}\hat{u}_{tj},$$

  where $\hat{u}$'s are the OLS residuals.

- Finally, note when $X_i = \underline{X}$, for all $i = 1 \ldots m$ equations, the GLS and OLS estimators are the same. Ie, when all X variables in each of the m equations are the same, the GLS and OLS estimators are equivalent. To see this note that $X = I_T \otimes \underline{X}$.

10 Lecture 10: October 4, 2005

10.1 Panel Data Models

Introduction

- These are also called longitudinal data sets or cross-section time-series. Consider the model:
$$y_{ti} = \alpha_{ti} + x_{ti1}\beta_{ti1} + \cdots + x_{tik}\beta_{tiK} + v_{ti}.$$
Or:
$$y_{ti} = \alpha_{ti} + \sum_{k=1}^{K} x_{tik}\beta_{tik} + v_{ti}.$$

- Since we have more parameters than observations, we need to make some simplifying assumptions in order to identify the system. Note i indexes an individual (say) and t indexes time.

- For example, consider 2 individuals and T periods. Suppose there is only one X variable. Then:
$$i = 1 \begin{cases} y_{11} = \alpha_{11} + x_{11}\beta_{11} + v_{11} \\ y_{21} = \alpha_{21} + x_{21}\beta_{21} + v_{21} \\ \vdots \\ y_{T1} = \alpha_{T1} + x_{T1}\beta_{T1} + v_{T1} \end{cases}$$
$$i = 2 \begin{cases} y_{12} = \alpha_{12} + x_{12}\beta_{12} + v_{12} \\ y_{22} = \alpha_{22} + x_{22}\beta_{22} + v_{22} \\ \vdots \\ y_{T2} = \alpha_{T2} + x_{T2}\beta_{T2} + v_{T2} \end{cases}$$

- So, suppose you have the system:
$$y_{ti} = \alpha + x_{ti}\beta + z_{ti}\gamma + v_{ti}, \quad t = 1 \ldots T, \; i = 1 \ldots N.$$
But suppose the z's are unobserved. If you just run y on x, you'll get a biased estimate of β. Then consider the model:
$$y_{ti} = \alpha_i + x_{ti}\beta + v_{ti}, \quad \alpha_i = \alpha + z_i\gamma.$$
If we can decompose like this, we'll get a consistent estimate of β, our parameter of interest. We just have to control for the effects of the z's (control for the fixed effects).

Fixed / Random Effects Model

- So, again, consider the model:
$$y_{ti} = \alpha_{ti} + \beta_{ti}x_{ti} + v_{ti}.$$

As it stands, we have NxT observations and $2xNxT$ parameters to estimate. So suppose:

$$\alpha_{ti} = \alpha + \mu_i + \lambda_t.$$

So we posit that the intercept depends on both an individual specific effect and a time effect. Also assume $\beta_{ti} = \beta \; \forall \; t, i$. Now we only have $2+N+T$ parameters to estimate.

- So for $N = 2$, the system looks like:

$$i = 1 \begin{cases} y_{11} = \alpha + \mu_1 + \lambda_1 + x_{11}\beta + v_{11} \\ y_{21} = \alpha + \mu_1 + \lambda_2 + x_{21}\beta + v_{21} \\ \vdots \\ y_{T1} = \alpha + \mu_1 + \lambda_T + x_{T1}\beta + v_{T1} \end{cases}$$

$$i = 2 \begin{cases} y_{12} = \alpha + \mu_2 + \lambda_1 + x_{12}\beta + v_{12} \\ y_{22} = \alpha + \mu_2 + \lambda_2 + x_{22}\beta + v_{22} \\ \vdots \\ y_{T2} = \alpha + \mu_2 + \lambda_T + x_{T2}\beta + v_{T2} \end{cases}$$

- For a general system of NxT observations, a vector of α's could be written: αe_{NT}, the matrix of μ's can be written $(I_N \otimes e_T)\mu$, and the time effects can be written $(e_N \otimes I_T)\lambda$, so the entire system is:

$$y = e_{NT}\alpha + (I_N \otimes e_T)\mu + (e_N \otimes I_T)\lambda + X\beta + v,$$

where y is an $NTx1$ vector.

- We don't know what α, μ and λ are, so we need to transform the data in some way to make those terms go to zero and thus the model is identified and we get a unbiased estimate of β.

Some Useful Matrices

- Define:

$$A_{NT} = (I_N \otimes e_T)(I_N \otimes e_T') = I_N \otimes e_T e_T'.$$

$$B_{NT} = (e_N \otimes I_T)(e_N' \otimes I_T) = e_N e_N' \otimes I_T.$$

$$J_{NT} = (e_N \otimes e_T)(e_N' \otimes e_T') = (e_T \otimes e_N)(e_T' \otimes e_N') = e_{NT}' e_{NT}.$$

- Thus J_{NT} is a square $NTxNT$ matrix of ones. B_{NT} is a square $NTxNT$ matrix with I_T in every position (NxN times). And A_{NT} is a square $NTxNT$ matrix with a TxT matrix of ones in each diagonal entry and TxT matrices of zeros everywhere else.

- So to motivate the next section, recall the normal equations from OLS:

$$(X'X)\beta = X'y,$$

where we have a constant: $x_{t1} = 1$. Then $\hat{\beta}' = [\hat{\beta}_1, \underline{\hat{\beta}'}]$ and:

$$\underline{\hat{\beta}} = (X_*' X_*')^{-1} X_*' y_*,$$

with:

$$X_* = (I_T - \frac{e_T e_T'}{T})\underline{X},$$

or the deviations from the mean. We're going to do a similar thing next with the panel data model.

- Consider the following means across time, individuals, and overall:

$$\text{Mean Across Time: } \bar{z}_{\cdot i} = \frac{1}{T}\sum_{t=1}^{T} z_{ti}.$$

$$\text{Mean Across Individuals: } \bar{z}_{t\cdot} = \frac{1}{N}\sum_{i=1}^{N} z_{ti}.$$

$$\text{Mean Across Time and Individuals: } \bar{z}_{\cdot\cdot} = \frac{1}{NT}\sum_{t=1}^{T}\sum_{i=1}^{N} z_{ti}.$$

- So:

$$\frac{A_{NT}}{T} z = \begin{bmatrix} \bar{z}_{\cdot 1} \\ \bar{z}_{\cdot 2} \\ \vdots \\ \bar{z}_{\cdot N} \end{bmatrix} \otimes e_T.$$

And,

$$\frac{B_{NT}}{T} z = e_N \otimes \begin{bmatrix} \bar{z}_{1\cdot} \\ \bar{z}_{2\cdot} \\ \vdots \\ \bar{z}_{T\cdot} \end{bmatrix}.$$

And,

$$\frac{J_{NT}}{NT} z = e_{NT} \bar{z}_{\cdot\cdot}.$$

- So the key to all of this are these three equations. The i^{th} element of …

$$(I_{NT} - \frac{A_{NT}}{T})z \text{ is } z_{ti} - \bar{z}_{\cdot i}$$

$$(I_{NT} - \frac{B_{NT}}{T})z \text{ is } z_{ti} - \bar{z}_{t\cdot}$$

$$(I_{NT} - \frac{J_{NT}}{T})z \text{ is } z_{ti} - \bar{z}_{\cdot\cdot}$$

So these are all deviations from the means (by time, individual household, and overall).

Some Other Useful Matrices

- Define:

$$Q_0 = I_{NT} - \frac{A_{NT}}{T} - \frac{B_{NT}}{N} + \frac{J_{NT}}{NT}, \;\; \overline{Q}_0 = \overline{\overline{Q}}_0 = I_{NT} - \frac{A_{NT}}{T}.$$

$$Q_1 = \overline{Q}_1 = \frac{A_{NT}}{T} - \frac{J_{NT}}{NT}, \;\; \overline{\overline{Q}}_1 = \frac{A_{NT}}{T}.$$

$$Q_2 = \frac{B_{NT}}{N} - \frac{J_{NT}}{NT}.$$

$$Q_3 = \frac{J_{NT}}{NT}.$$

- Finally, we show that Q_1 is idempotent. Note:

$$Q_1 = \frac{A_{NT}}{T} - \frac{J_{NT}}{NT}.$$

Consider the following:

$$\begin{aligned} A_{NT}A_{NT} &= (I_N \otimes e_T e_T')(I_N \otimes e_T e_T') \\ &= I_N \otimes e_T e_T' e_T e_T' \\ &= T[I_N \otimes e_T e_T'] \\ &= TA_{NT} \end{aligned}$$

$$\begin{aligned} J_{NT}J_{NT} &= e_{NT}e_{NT}'e_{NT}e_{NT}' \\ &= NT[e_{NT}e_{NT}'] \\ &= NTJ_{NT} \end{aligned}$$

$$\begin{aligned} A_{NT}J_{NT} &= (I_N \otimes e_T e_T')(e_N e_N' \otimes e_T e_T') \\ &= e_N e_N' \otimes e_T e_T' e_T e_T' \\ &= T[e_N e_N' \otimes e_T e_T'] \\ &= TJ_{NT} \end{aligned}$$

So,

$$\begin{aligned}
Q_1Q_1 &= (\frac{A_{NT}}{T} - \frac{J_{NT}}{NT})(\frac{A_{NT}}{T} - \frac{J_{NT}}{NT}) \\
&= \frac{A_{NT}A_{NT}}{T^2} - \frac{2A_{NT}J_{NT}}{NT^2} + \frac{J_{NT}J_{NT}}{N^2T^2} \\
&= \frac{TA_{NT}}{T^2} - \frac{2TJ_{NT}}{NT^2} + \frac{NT * J_{NT}}{N^2T^2} \\
&= \frac{A_{NT}}{T} - \frac{2J_{NT}}{NT} + \frac{J_{NT}}{NT} \\
&= \frac{A_{NT}}{T} - \frac{J_{NT}}{NT} \\
&= Q_1
\end{aligned}$$

Idempotent.

11 Lecture 11: October 6, 2005

11.1 Panel Data Models

Review from Last Lecture

- Given our useful matrices we defined last time, we can use them to find various deviations from sample means as follows (typical element listed):

$$(I_{NT} - \frac{J_{NT}}{NT})y \leftrightsquigarrow y_{ti} - \overline{y}_{..}$$

$$(I_{NT} - \frac{A_{NT}}{T})y \leftrightsquigarrow y_{ti} - \overline{y}_{.i}$$

$$(I_{NT} - \frac{B_{NT}}{N})y \leftrightsquigarrow y_{ti} - \overline{y}_{t.}$$

- We also defined another matrix:

$$Q_0 = I_{NT} - \frac{A_{NT}}{T} - \frac{B_{NT}}{N} + \frac{J_{NT}}{NT}.$$

What does this do to our vector of y's :

$$Q_0 y = y_{ti} - \overline{y}_{.i} - \overline{y}_{t.} + \overline{y}_{..}.$$

Or,

$$Q_0 y = \underbrace{y_{ti} - \overline{y}_{..}}_{overall\ variation} - \underbrace{(\overline{y}_{.i} - \overline{y}_{..})}_{individ\ spec\ var} - \underbrace{(\overline{y}_{t.} - \overline{y}_{..})}_{time\ spec\ var}.$$

- We also have Q_1, Q_2, and Q_3 which subtract other sample means. All the Q matrices are symmetric, idempotent, and orthogonal to each other.

- **Lemma M1** There are many parts to this, but suppose we want to prove that $Q_1 e_{NT} = 0$.

Proof:

$$\begin{aligned}
Q_1 e_{NT} &= \left(\frac{A_{NT}}{T} - \frac{J_{NT}}{NT} \right) e_{NT} \\
&= \left(\frac{I_N \otimes e_T e_T'}{T} - \frac{e_{NT} e_{NT}'}{NT} \right) e_{NT} \\
&= \left(I_N \otimes \frac{e_T e_T'}{T} - \frac{e_{NT} e_{NT}'}{NT} \right) e_{NT} \\
&= (I_N \otimes \frac{e_T e_T'}{T}) e_{NT} - \frac{e_{NT} e_{NT}'}{NT} e_{NT} \\
&= (I_N \otimes \frac{e_T e_T'}{T})(e_N \otimes e_T) - \frac{e_{NT} e_{NT}' e_{NT}}{NT} \\
&= (e_N \otimes \frac{e_T e_T' e_T}{T}) - \frac{e_{NT} NT}{NT} \\
&= (e_N \otimes \frac{e_T T}{T}) - \frac{e_{NT} NT}{NT} \\
&= (e_N \otimes e_T) - e_{NT} \\
&= e_{NT} - e_{NT} = 0
\end{aligned}$$

QED.

- **Lemma M3** The variance/covariance matrix of our estimator will look like:

$$\Omega = \sum_{k=0}^{3} \sigma_k^2 Q_k,$$

and this lemma says:

$$\Omega^{-1} = \sum_{k=0}^{3} \sigma_k^{-2} Q_k.$$

This also holds more generally when the σ_k's are matrices.

- He jumped ahead at this point and mentioned that on page 43 of his notes the variance/covariance matrix will be something like:

$$\Omega = \sigma_\mu^2 A_{NT} + \sigma_\lambda^2 B_{NT} + \sigma_v^2 I_{NT}.$$

If we take this and add and subtract correctly (see notes), we get Ω in terms of the Q's. We can then apply lemma M3 to invert the matrix and we can do GLS.

Single Equation Regression Model: Two Way Error Components

- Consider the following model:

$$y_{ti} = \alpha_{ti} + \sum_{k=1}^{K} x_{tik}\beta_k + v_{ti}, \quad t = 1 \ldots T, i = 1 \ldots N.$$

  And assume:

$$\alpha_{ti} = \alpha + \mu_i + \lambda_t.$$

  Since we have too many coefficients to estimate in the original model, we assume this structure on the intercept term so we can identify the model. The μ parameters will capture the variation across individuals and the λ parameters will capture variation across time. We have assumed that the β's are the same across individuals and time. We could relax this (in the case of a "random coefficients model"), but this makes it much more difficult to invert Ω. Since there is an error component for the individuals and for time, we call this a two way error component model.

- If we know the μ's and λ's are uncorrelated with the x's, we can just assume they are part of the error term and estimate via OLS consistently. If we can't assume that, we have a fixed effect type model which we estimate with a dummy variable estimator (next). Our model becomes:

$$y_{ti} = \alpha + \mu_i + \lambda_t + \sum_{k=1}^{K} x_{tik}\beta_k + v_{ti}, \quad t = 1 \ldots T, i = 1 \ldots N.$$

- We can write the model in matrix form (see last lecture) as:

$$y = e_{NT}\alpha + (I_N \otimes e_T)\mu + (e_N \otimes I_T)\lambda + X\beta + v.$$

  We assume that:

$$v_{ti} \sim iid(0, \sigma_v^2), \quad \sigma_v^2 < \infty.$$

  Or,

$$E[v] = 0, \quad E[vv'] = \sigma_v^2 I_{NT}.$$

  We also assume that all the matrices of demeaned data are non-singular (ie, they converge to non-singular matrices as T and N go to infinity together).

- So the problem now with our model is we have perfect multicollinearity among the constant term, the μ's and λ's. So we impose two restrictions:
  - (1) $\sum_{i=1}^{N} \mu_i = 0$.
  - (2) $\sum_{t=1}^{T} \lambda_t = 0$.

- Since we assume a "Fixed Effects Model," the μ_i's and λ_t's are assumed to be fixed and unknown parameters. If we assumed them to be random (and possibly correlated with the x's), we would have a "Random Effects Model" and would have to estimate with GLS.

- Given these restrictions, our model becomes:

$$y = W\gamma + X\beta + v,$$

with,

$$\gamma = (\alpha, \mu_1, \ldots \mu_{N-1}, \lambda_1, \ldots, \lambda_{T-1})'.$$

- The normal equations are thus:

$$W'y = W'W\hat{\gamma} + W'X\hat{\beta},$$

$$X'y = X'W\hat{\gamma} + X'X\hat{\beta}.$$

Which induces the following Least Squares Dummy Variable estimator:

$$\hat{\beta}_{LSDV} = (X'Q_0X)^{-1}X'Q_0y.$$

Also called the fixed effects or within estimator.

- To actually run a model like this, transform the data using Q_0 and run OLS on the transformed model:

$$Q_0y = Q_0X\beta + Q_0v.$$

- Once we have $\hat{\beta}$, we can estimate the other parameters as well

$$\hat{\alpha}_{LSDV} = \overline{y}_{\cdot\cdot} - \sum_{k=1}^{K} \overline{x}_{\cdot\cdot k}\hat{\beta}_{LSDV,k}$$

$$\hat{\mu}_{LSDV,i} = \overline{y}_{\cdot i} - \hat{\alpha}_{LSDV} - \sum_{k=1}^{K} \overline{x}_{\cdot ik}\hat{\beta}_{LSDV,k}$$

$$\hat{\lambda}_{LSDV,i} = \overline{y}_{t\cdot} - \hat{\alpha}_{LSDV} - \sum_{k=1}^{K} \overline{x}_{t\cdot k}\hat{\beta}_{LSDV,k}$$

- From Prucha's notes: We note that the within estimator cannot estimate the effects of any time invariant and unit-invariant variables, since the Q_0-transformation wipes out all those variables.

12 Lecture 12: October 11, 2005

12.1 Panel Data Models

Asymptotic Properties of the LSDV Estimator

- Recall our model from last time:

$$y = e_{NT}\alpha + (I_N \otimes e_T)\mu + (e_N \otimes I_T)\lambda + X\beta + v.$$

And our estimator we formed:

$$\hat{\beta}_{LSDV} = (X'Q_0X)^{-1}X'Q_0y = \beta + (X'Q_0X)^{-1}X'Q_0v.$$

Transforming by Q_0 wipes out the time and individual specific effects.

- Then the asymptotics are:

$$\sqrt{NT}(\hat{\beta}_{LSDV} - \beta) = \left(\frac{1}{NT}X'Q_0X\right)^{-1}\frac{1}{\sqrt{NT}}X'Q_0v \to^d N(0, lim\ \sigma_v^2(\frac{1}{NT}X'Q_0X)^{-1}),$$

because $v \sim iid(0, \sigma_v^2)$.

- Clearly $\hat{\beta}_{LSDV}$ is unbiased. Also,

$$Var(\hat{\beta}_{LSDV}) = \sigma_v^2(X'Q_0X)^{-1} = \frac{1}{NT}\sigma_v^2(\frac{1}{NT}X'Q_0X)^{-1} \to 0,$$

so by chebychev, $\hat{\beta}_{LSDV}$ is also consisent.

- Consider the residuals:

$$\hat{v} = y - e_{NT}\hat{\alpha} - (I_N \otimes e_T)\hat{\mu} - (e_N \otimes I_T)\hat{\lambda} - X\hat{\beta}.$$

Then:

$$\hat{\sigma}_v^2 = \frac{\hat{v}'\hat{v}}{(N-1)(T-1)-K}.$$

Testing for Fixed Effects

- Since the fixed effects model is just a dummy variable model, we can test for fixed effects by doing an F test on the model with the restrictions that all the μ's and λ's are zero and then on the full model. Null:

$$\mu_1 = \cdots = \mu_{N-1} = \lambda_1 = \cdots = \lambda_{T-1} = 0.$$

Statistic:

$$F = \frac{(ESS_R - ESS_U)/(N+T-2)}{ESS_U/[(N-1)(T-1)-K]} \to^d \frac{\chi^2(N+T-2)}{N+T-2}.$$

Random Effects Model

- Now consider the model:

$$y = e_{NT}\alpha + X\beta + u,$$

with:

$$u = (I_N \otimes e_T)\mu + (e_N \otimes I_T)\lambda + v.$$

So we include the time and individual effects in the error term. The μ's and λ's are assumed to be iid conditional on the X's. They do NOT depend on the X's.

- We also assume,

$$\mu_i \sim iid(0, \sigma_\mu^2), \quad \lambda_t \sim iid(0, \sigma_\lambda^2), \quad v_{ti} \sim iid(0, \sigma_v^2).$$

And these variances are also independent of each other so $E[\mu v'] = E[\lambda v'] = E[\mu\lambda'] = 0$.

- Thus $E[u] = 0$ and,

$$E[uu'] = \Omega = [\cdot][\cdot]' = \sigma_\mu^2 A_{NT} + \sigma_\lambda^2 B_{NT} + \sigma_v^2 I_{NT}.$$

This means:

$$E[u_{ti}u_{sj}] = \begin{cases} \sigma_\mu^2 + \sigma_\lambda^2 + \sigma_v^2, & i = j, t = s \\ \sigma_\mu^2, & i = j, t \neq s \\ \sigma_\lambda^2, & i \neq j, t = s \\ 0, & else \end{cases}$$

- And as we have shown in the past:

$$\Omega = \sigma_v^2 Q_0 + \sigma_1^2 Q_1 + \sigma_2^2 Q_2 + \sigma_3^2 Q_3,$$

with:

$$\sigma_1^2 = T\sigma_\mu^2 + \sigma_v^2,$$
$$\sigma_2^2 = N\sigma_\lambda^2 + \sigma_v^2,$$
$$\sigma_3^2 = T\sigma_\mu^2 + N\sigma_\lambda^2 + \sigma_v^2.$$

- And, again as we have shown:

$$\Omega^{-1} = \sigma_v^{-2} Q_0 + \sigma_1^{-2} Q_1 + \sigma_2^{-2} Q_2 + \sigma_3^{-2} Q_3,$$

So this is our variance/covariance matrix but we don't usually know all these variance terms. We will estimate them soon.

- Then if we write the model:

$$y = e_{NT}\alpha + X\beta + u = Z\delta + u,$$

with $\delta = [\alpha, \beta']'$, our BLUE estimator is:

$$\hat{\delta}_{GLS} = (Z'\Omega^{-1}Z)^{-1}Z'\Omega^{-1}y.$$

- **Remark** In order to make a comparison with the fixed effects model, we can write:

  $$\hat{\beta}_{GLS} = (X'\Phi^{-1}X)^{-1}X'\Phi^{-1}y,$$

  with:

  $$\Phi^{-1} = \sigma_v^{-2}Q_0 + \sigma_1^{-2}Q_1 + \sigma_2^{-2}Q_2.$$

  See Prucha's notes. The difference between this and the fixed effects model is that we now have different weights when we transform the data depending on the variances of the μ's and λ's. This breakdown was only done to show the difference in the weightings. In practice, computing δ above (including the intercept) will get you the β parameters as well.

- For Feasible GLS, we need estimators for the variances which induces an estimated variance/covariance matrix:

  $$\hat{\sigma}_\mu^2, \hat{\sigma}_\lambda^2, \hat{\sigma}_v^2 \Longrightarrow \hat{\Omega}^{-1},$$

  and thus,

  $$\hat{\delta}_{FGLS} = (Z'\hat{\Omega}^{-1}Z)^{-1}Z'\hat{\Omega}^{-1}y.$$

Interpretation as Weighted Average of Within and Between Estimators

- Consider again the model:

  $$y = e_{NT}\alpha + X\beta + (I_N \otimes e_T)\mu + (e_N \otimes I_T)\lambda + v.$$

  If we transform by Q_0, we get the fixed effects or "within estimator". If we tranform by Q_1 or Q_2, we get two different "between estimators."

- If you want, you can interpret the random effects estimator, $\hat{\beta}_{GLS}$, as the weighted average of the within estimator and the two between estimators as shown in the notes. The idea is that if the σ's are too large, the random effects estimator tends to the fixed effects estimator as using the variances really isn't buying you much and you might as well run a fixed effects model. If the σ's are small, there may be gains from estimating the random effects model.

13 Lecture 13: October 13, 2005

13.1 Panel Data Models - More on Random Effects Model

- Recall our random effects model:

$$y = e_{NT}\alpha + X\beta + u,$$

with:

$$u = (I_N \otimes e_T)\mu + (e_N \otimes I_T)\lambda + v.$$

OLS Estimator

- Consider the OLS estimator of β:

$$\hat{\beta}_{OLS} = [X'(I_{NT} - \frac{e_{NT}e'_{NT}}{NT})X]^{-1}X'(I_{NT} - \frac{e_{NT}e'_{NT}}{NT})y,$$

or,

$$\hat{\beta}_{OLS} = \beta + [X'(I_{NT} - \frac{e_{NT}e'_{NT}}{NT})X]^{-1}X'(I_{NT} - \frac{e_{NT}e'_{NT}}{NT})u.$$

So we are just demeaning all our data and we get the OLS estimator. Since $E[u|X] = 0$, $\hat{\beta}_{OLS}$ is unbiased. Consider the variances:

$$VC(u) = \sigma_v^2 Q_0 + \sigma_1^2 Q_1 + \sigma_2^2 Q_2 + \sigma_3^2 Q_3,$$

so,

$$VC(\hat{\beta}_{OLS}) = \frac{1}{(NT)^2} X'(\sum_{i=0}^{3} \sigma_i^2 Q_i)X.$$

So since the variance goes to zero, $\hat{\beta}_{OLS}$ is also consistent.

LSDV Estimator

- Consider the LSDV estimator of β:

$$\hat{\beta}_{LSDV} = (X'Q_0X)^{-1}X'Q_0y,$$

or,

$$\hat{\beta}_{LSDV} = \beta + (\frac{1}{NT}X'Q_0X)^{-1}\frac{1}{NT}X'Q_0v.$$

Clearly $\hat{\beta}_{LSDV}$ is unbiased and since the variance again goes to zero, consistent. Note the asymptotic distribution:

$$\sqrt{NT}(\hat{\beta}_{LSDV} - \beta) = (\frac{1}{NT}X'Q_0X)^{-1}\frac{1}{\sqrt{NT}}X'Q_0v \rightarrow^d N(0, lim\ \sigma_v^2(\frac{1}{NT}X'Q_0X)^{-1}).$$

GLS Estimator

- Finally, if the random effects assumptions hold, then the GLS estimator is the most efficient estimator of β:

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

Again, $E[\hat{\beta}_{GLS}] = \beta$ and $plim(\hat{\beta}_{GLS}) = \beta$. Note the asymptotic distribution:

$$\sqrt{NT}(\hat{\beta}_{GLS} - \beta) \rightarrow^d N(0, lim\ \sigma_v^2(\frac{1}{NT}X'Q_0X)^{-1}).$$

This is the SAME as the LSDV estimator!! Thus the asymptotic distributions of the GLS and LSDV estimators are the same so if you have a large enough sample, just run a fixed effects model using LSDV and everything will be fine.

Feasible GLS Estimator for the Random Effects Model

- **Theorem** As long as:

$$plim\ \hat{\sigma}_v^2 = \sigma_v^2,\ \ plim\ \hat{\sigma}_\mu^2 = \sigma_\mu^2,\ \ plim\ \hat{\sigma}_\lambda^2 = \sigma_\lambda^2,$$

then:

$$plim\ \sqrt{NT}(\hat{\beta}_{FGLS} - \hat{\beta}_{GLS}) = 0.$$

So as long as you have consistent estimators for the variances, the asymptotic distributions of the FGLS and the true GLS estimators will be the same. Since OLS and LSDV both produce consistent estimates, we can use either method.

Estimating Variances

- So how are we going to estimate these variance terms? Recall:

$$VC(u) = E[uu'] = \sigma_v^2 Q_0 + \sigma_1^2 Q_1 + \sigma_2^2 Q_2 + \sigma_3^2 Q_3.$$

Now consider $Q_0 u$. Clearly,

$$E[Q_0 u] = 0,$$

and,

$$VC(Q_0 u) = E[Q_0 uu' Q_0] = Q_0 E[uu'] Q_0 = \sigma_v^2 Q_0.$$

This is because all the Q matrices are orthogonal to each other so only the first term doesn't get wiped out. Recall as well that Q_0 is idempotent.

- So let $u_* = Q_0 u$ and consider $u_*' u_* = u' Q_0 u$, a quadratic form. Then:

$$\begin{aligned} E[u_*' u_*] &= E[u' Q_0 u] \\ &= E[tr(u' Q_0 u)] \\ &= E[tr(Q_0 u u')] \\ &= tr[E(Q_0 u u')] \\ &= tr[Q_0 E(u u')] \\ &= tr[\sigma_v^2 Q_0] \\ &= \sigma_v^2 tr[Q_0] \\ &= \sigma_v^2 (NT - T - N + 1) \end{aligned}$$

  So,

$$E\Big[\frac{u_*' u_*}{NT - T - N + 1}\Big] = E\Big[\frac{u' Q_0 u}{NT - T - N + 1}\Big] = \sigma_v^2.$$

  Thus we form our Analysis of Variance (AOV) estimator:

$$\tilde{\sigma}_v^2 = \frac{u' Q_0 u}{NT - T - N + 1}.$$

  Similarly, we have:

$$\tilde{\sigma}_1^2 = \frac{u' Q_1 u}{N - 1}, \quad \tilde{\sigma}_2^2 = \frac{u' Q_2 u}{T - 1}.$$

  Also, since $\sigma_1^2 = T\sigma_\mu^2 + \sigma_v^2$ and $\sigma_2^2 = N\sigma_\lambda^2 + \sigma_v^2$, then:

$$\tilde{\sigma}_\mu^2 = \frac{\tilde{\sigma}_1^2 - \tilde{\sigma}_v^2}{T}.$$

$$\tilde{\sigma}_\lambda^2 = \frac{\tilde{\sigma}_2^2 - \tilde{\sigma}_v^2}{N}.$$

- **Lemma** Let $\tilde{u} = y - e_{NT}\tilde{\alpha} - X\tilde{\beta}$ be an estimator of the disturbances based on consistent parameter estimators. Then the AOV estimators above are consistent. Thus the residuals from OLS or LSDV will be fine for estimating our variances.

Testing for Random Effects

- Consider the null hypothesis:

$$H_0 : \mu_1 = \cdots = \mu_n = \lambda_1 = \cdots = \lambda_T = 0.$$

  This corresponds, in the random effects model setting, to:

$$H_0 : \sigma_\mu^2 = \sigma_\lambda^2 = 0,$$

because if the means are already all zero, if the variances are both zero, then we can safely reject random effects.

- To run this test, we use a Breusch Pagan LM test as shown in Prucha's notes. We compute the estimator under the null which correponds to the OLS estimator.

Fixed versus Random Effects

- In the Random effects setting, we believe the μ's, for example, are not scattered completely randomly, but rather we think they might be bunched around some mean value. If $\sigma_\mu^2 \to \infty$, then the random effects estimator $\to$ the fixed effects estimator.
- In SMALL samples, if the random effects assumptions hold, run GLS because there may be gains in efficiency to running the random effects model. As the sample size grows, you might as well run a fixed effects model because the efficiency gains vanish.
- Important caveat: If the random effects assumptions do NOT hold but you still estimate a random effects model, the random effects (GLS) estimator is inconsistent. We need a test to figure out if this is the case. Bring in Hausman.

Hausman Test

- Consider a null and alternative hypotheses, H_0 and H_1, which are possibly not parametric (eg, H_0: X and u are uncorrelated).
- Consider two estimators:

$$\hat{\theta}_T : \;\; plim\ \hat{\theta}_T = \theta, \text{ efficient, under } H_0.$$

$$\tilde{\theta}_T : \;\; plim\ \tilde{\theta}_T = \theta, \text{ under } H_0 \text{ and } H_1. \text{ But inefficient under } H_0.$$

Thus $plim\ \hat{\theta}_T \neq \theta$ under H_1.

- So if $\hat{\theta}_T - \tilde{\theta}_T$ is big, then we might suspect that the null does not hold and H_1 is more likely to be true.
- Consider the following Hausman test statistic:

$$(\underbrace{\hat{\theta}_T}_{px1} - \tilde{\theta}_T)'[VC(\hat{\theta}_T - \tilde{\theta}_T)]^{-1}(\hat{\theta}_T - \tilde{\theta}_T) \to \chi^2(p).$$

Note, under the above assumptions, we can estimate the variance of the difference as

$$VC(\hat{\theta}_T - \tilde{\theta}_T) = VC(\tilde{\theta}_T) - VC(\hat{\theta}_T).$$

We don't need to know the covariance structure.

14 Lecture 14: October 18, 2005

14.1 More Panel Data Models with Endogenous Effects

Mundlak Model

- He we assume a one-way error component model as usual:

$$y_{ti} = \alpha + \mu_i + \sum_{k=1}^{K} x_{tik}\beta_k + v_{ti},$$

but now we introduce some structure to the correlation of the μ's and the X's:

$$\mu_i = \sum_{k=1}^{K} \overline{x}_{\cdot ik}\pi_i + \epsilon_i.$$

Thus Mudlak assumes that the individual effects can be expressed as a linear function of the time averages of all the explanatory variables.

- So we want to estimate α, the β's and the π's.
- Write the model in matrix notation:

$$y = e_{NT}\alpha + (I_N \otimes e_T)\mu + X\beta + v,$$

$$\mu = (I_N \otimes \frac{e'_T}{T})X\pi + \epsilon.$$

Or,

$$y = e_{NT}\alpha + \frac{A_{NT}}{T}X\pi + X\beta + (I_N \otimes e_T)\epsilon + v.$$

- Recall that $\overline{Q}_0 = \overline{\overline{Q}}_0 = I_{NT} - \dfrac{A_{NT}}{T}$, and $\overline{\overline{Q}}_1 = \dfrac{A_{NT}}{T}$.
- Consider our disturbance term: $u = (I_N \otimes e_T)\epsilon + v$. It is shown in the notes:

$$VC(u) = \sigma_\epsilon^2 A_{NT} + \sigma_v^2 I_{NT} = \sigma_v^2\overline{\overline{Q}}_0 + \sigma_*^2\overline{\overline{Q}}_1,$$

with,

$$\sigma_*^2 = T\sigma_\epsilon^2 + \sigma_v^2.$$

- Then our GLS estimators become:

$$\tilde{\beta}_{GLS} = \hat{\beta}_{LSDV} = (X'\overline{\overline{Q}}_0X)^{-1}X'\overline{\overline{Q}}_0y,$$

$$\tilde{\pi}_{GLS} = \tilde{\beta}_B - \hat{\beta}_{LSDV},$$

$$\tilde{\alpha}_{GLS} = \overline{y} - \overline{x}\tilde{\beta}_B,$$

where $\tilde{\beta}_B = (X'\overline{\overline{Q}}_1 X)^{-1} X'\overline{\overline{Q}}_1 y$. Thus the GLS estimator in this case is just the FIXED effects estimator.

- Mundlak said this is reason enough to never use the random effects model and always use fixed effects. However, as we have said before, if the assumptions of random effects hold, there are efficiency gains to running the random effects model.

Hausman Taylor Model

- The Mundlak (1978) specification assumes that a priori all explanatory variables are potentially correlated with the individual effects. The random effects specification assumes a priori that none of the explanatory variables are correlated with the regressors. The Hausman and Taylor (1981) specification now allows for some of the explanatory variables to be correlated with the regressors and distinguishes explicitly between regressors that vary over units and time and those that only vary over units.

- Consider the following model:

$$y_{ti} = \sum_{k=1}^{K} x_{tik}\beta_k + \sum_{l=1}^{L} z_{il}\gamma_l + u_{ti},$$

with,

$$u_{ti} = \mu_i + v_{ti}.$$

So the z's do not vary with time.

- In matrix notation:

$$y = X\beta + Z\gamma + u, \quad u = (I_N \otimes e_T)\mu + v.$$

- Consider decomposing the X and Z variables as follows:

$$X = (X_1, X_2), \quad Z = (Z_1, Z_2),$$

such that:

$$v \perp Z, X, \text{ and } \mu \perp X_1, Z_1.$$

So X and Z are independent of v, but only X_1 and Z_1 are independent of μ. However X_2 and Z_2 may be correlated with the individual effects.

- Given that $E[u|X_1, Z_1] = 0$, we have moment conditions: $E[X_1'u] = 0$ and $E[Z_1'u] = 0$. Our variance/covariance matrix can be written:

$$E[uu'] = \Omega = E[uu'|X_1, Z_1] = \sigma_v^2\overline{\overline{Q}}_0 + \sigma_1^2\overline{\overline{Q}}_1.$$

- See the notes for an aside on GMM. The idea is that using the two moment conditions above, we can define our GMM estimator which is really an IV or 2SLS estimator.

- Since X_1 and Z_1 are independent of μ, we can use them as instruments for Z_2 and X_2. However, in general we need at least as many instruments as we had original variables, so we need more. Hausman Taylor consider the following set of instruments:

$$H = [\overline{\overline{Q}}_0 X, \overline{\overline{Q}}_1 X_1, Z_1] = [\overline{\overline{Q}}_0 X_1, \overline{\overline{Q}}_0 X_2, \overline{\overline{Q}}_1 X_1, Z_1].$$

It's clear that X_1 can be used as instruments, but why are $\overline{\overline{Q}}_0 X_2$ good instruments? Consider:

$$Cov(\overline{\overline{Q}}_0 X_2, \Omega^{-1/2} u) = 0,$$

so because we're creating deviations from the mean (not linearly dependent), we can use them as instruments.

- So consider transforming the data as follows:

$$y^* = \sigma_v \Omega^{-1/2} y = [I_{NT} - (1-\theta)\overline{\overline{Q}}_1] y$$

$$X^* = \sigma_v \Omega^{-1/2} X = [I_{NT} - (1-\theta)\overline{\overline{Q}}_1] X$$

$$Z^* = \sigma_v \Omega^{-1/2} Z = [I_{NT} - (1-\theta)\overline{\overline{Q}}_1] Z = \theta Z$$

where $\theta = \dfrac{\sigma_v}{\sigma_1}$

- Then the IV or 2SLS estimator corresponding to instruments H with $P_H = H(H'H)^{-1}H'$ are:

$$\begin{bmatrix} \hat{\beta}_{HT} \\ \hat{\gamma}_{HT} \end{bmatrix} = \left\{ \begin{bmatrix} X^{*\prime} \\ Z^{*\prime} \end{bmatrix} P_H \begin{bmatrix} X^* & Z^* \end{bmatrix} \right\}^{-1} \begin{bmatrix} X^{*\prime} \\ Z^{*\prime} \end{bmatrix} P_H y^*.$$

- So we need an estmate of θ. We do this in a few steps.
  - (1) First, transform the data according to:

  $$\overline{\overline{Q}}_0 y = \overline{\overline{Q}}_0 X\beta + \overline{\overline{Q}}_0 u,$$

  and run OLS yielding consistent estimates of $\tilde{\beta}$.
  - (2) Second, transform the data according to:

  $$\overline{\overline{Q}}_1 y - \overline{\overline{Q}}_1 X\hat{\beta} = \underbrace{\overline{\overline{Q}}_1 Z}_{Z} \gamma + \overline{\overline{Q}}_1 u,$$

  and run an IV regression using X_1 and Z_1 as instruments to get a consistent estimate of $\tilde{\gamma}$.
  - (3) Create residuals:

  $$\tilde{u} = y - X\tilde{\beta} - Z\tilde{\gamma}.$$

  - (4) Create AOV estimators:

  $$\tilde{\sigma}_1^2 = \frac{\tilde{u}'\overline{\overline{Q}}_1 \tilde{u}}{N},$$

$$\tilde{\sigma}_v^2 = \frac{\tilde{u}'\overline{\overline{Q}}_0\tilde{u}}{N(T-1)},$$

$$\tilde{\sigma}_\mu^2 = \frac{\tilde{\sigma}_1^2 - \tilde{\sigma}_v^2}{T}.$$

– (5) Finally, form our estimate of θ:

$$\tilde{\theta} = \frac{\tilde{\sigma}_v}{\tilde{\sigma}_1}.$$

15 Lecture 15: October 20, 2005

15.1 Univariate Linear Dynamic Models

Autoregressive Models with iid Errors

- Consider the model:

$$y_t = a_1 y_{t-1} + \cdots + a_p y_{t-p} + z_{t\cdot}\gamma + u_t.$$

  Here, the $z_{t\cdot}$ are exogenous variables (and possibly their lags).

- We make the following 4 assumptions:
  - (1) $u_t \sim iid(0, \sigma^2)$, $E[u_t^4] < \infty$.
  - (2) The exogenous Z's are uniformly bounded in absolute value.
  - (3) The following holds:

$$lim_{T\to\infty} \frac{1}{T-\tau} \sum_{t=1}^{T-\tau} z'_{t\cdot} z_{(t+\tau)\cdot}$$

    is finite for all τ and further for $\tau = 0$, $\frac{1}{T} Z'Z$ converges to something finite and non-singular.
  - (4) The roots of the characteristic polynomial:

$$\lambda^p - a_1 \lambda^{p-1} - \cdots - a_p = 0,$$

    are all less than one in absolute value, ie, $|\lambda_i| < 1 \; \forall \; i$. This ensures we have a stable system.

- **Remark** Let's consider assumption 4. Rewrite the model:

$$y_t = a_1 y_{t-1} + \cdots + a_p y_{t-p} + \underbrace{f_t}_{z_{t\cdot}\gamma + u_t}.$$

  So we have a p^{th} order difference equation with forcing variable, f_t. The general solution is the sum of the homogenous and particular solutions. Consider first the homogenous part (set the forcing term to zero):

$$y_t = a_1 y_{t-1} + \cdots + a_p y_{t-p}.$$

$$y_t - a_1 y_{t-1} - \cdots - a_p y_{t-p} = 0.$$

  Guess a solution of the form $y_t^h = A\lambda^t$. Substituting in (dividing through by A),

$$\lambda^t - a_1 \lambda^{t-1} - \cdots - a_p \lambda^{t-p} = 0.$$

Divide through by λ^{t-p}:

$$\lambda^p - a_1\lambda^{p-1} - \cdots - a_p\lambda^0 = 0.$$

This is our characteristic polynomial which in general has roots: $\lambda_1, \ldots, \lambda_p$.

- **Remark** Consider the simple model:

$$y_t = ay_{t-1} + u_t, \quad |a| < 1, u_t \sim iid(0, \sigma^2).$$

Homogenous part:

$$y_t - ay_{t-1} = 0.$$

Homogenous guess is $y_t^h = A\lambda^t$. Plug in:

$$A\lambda^t - aA\lambda^{t-1} = 0.$$

$$\lambda^t - a\lambda^{t-1} = 0.$$

$$\lambda - a = 0.$$

$$a = \lambda.$$

So we only have one root and it equals a. Thus our homogenous solution is:

$$y_t^h = A\lambda^t = Aa^t.$$

To find a particular solution, consider writing the model in lag operator form:

$$
\begin{aligned}
(1-aL)y_t &= u_t \\
y_t^p &= \frac{1}{1-aL}u_t \\
&= \sum_{i=0}^{\infty} a^i L^i u_t \\
&= \sum_{i=0}^{\infty} a^i u_{t-i} \\
& \text{sub back into original model} \\
y_t^p &= a y_{t-1}^p + u_t \\
\sum_{i=0}^{\infty} a^i u_{t-i} &= a\sum_{i=0}^{\infty} a^i u_{t-i-1} + u_t \\
&= \sum_{i=0}^{\infty} a^{i+1} u_{t-i-1} + u_t \\
&= \sum_{j=1}^{\infty} a^j u_{t-j} + u_t \\
&= \sum_{j=0}^{\infty} a^j u_{t-j} \\
\sum_{i=0}^{\infty} a^i u_{t-i} &= \sum_{j=0}^{\infty} a^j u_{t-j} \\
& \text{So it works!}
\end{aligned}
$$

Thus our general solution is:

$$y_t = y_t^h + y_t^p = Aa^t + \sum_{i=0}^{\infty} a^i u_{t-i}.$$

If $E[u_{t-1}] = 0$, then $E[y_t] = Aa^t$. So consider a few cases.

- (1) A stationary solution features: $E[y_t] = E[y_{t+1}] = c$ and $E[y_t y_{t+1}] = h(s)$. Thus, for our process, it is only stationary if $A = 0$. This is equivalent to saying the process started at $t = -\infty$.
- (2) We could also fix $y_0 = y_0^*$, but this puts a restriction on $A \neq 0$ and the solution would no longer be stationary.

- (3) For a general process, consider the variance:

$$Var[y_t] = E[(y_t - E(y_t))^2] = E[\sum_{i=0}^{\infty} a^i u_{t-i}] = \frac{\sigma_u^2}{1-a^2},$$

as we have shown in the section on $AR(1)$ error processes.

- (4) Finally if $a = 1$, the model becomes:

$$y_t = y_{t-1} + u_t,$$

a random walk. The characteristic polynomial is $\lambda - a = 0 \Rightarrow a = \lambda \Rightarrow \lambda = 1 \Rightarrow$ a unit root! Also,

$$lim_{T\to\infty} Var(y_T) = lim_{T\to\infty} T\sigma_u^2 = \infty.$$

So the process explodes.

- So back to our original model. Write the model in matrix form as:

$$y = X\beta + u,$$

where X contains the lags of y and the Z variables and $\beta = (a_1, \ldots, a_p, \gamma')'$.

- We could estimate via OLS as:

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

Is it unbiased? Consider:

$$E[\hat{\beta}_{OLS}] = \beta + E[(X'X)^{-1}X'u] = \beta + E_x[(X'X)^{-1}X'E(u|X)].$$

So can we go further? Consider the model:

$$y_t = a_1 y_{t-1} + u_t, \; t = 1 \ldots T.$$

The second error term is thus:

$$u_2 = y_2 - a_1 y_1,$$

which has expected value:

$$E[u_2|y_1 \ldots y_T] = y_2 - a_1 y_1 \neq 0.$$

So in general, $E[u|X] \neq 0$, so the OLS estimator for β is BIASED.

- **Lemma** By a CLT from Schonfeld (1971), under the assumptions above, we have:
  - (1) $T^{-1}X'X \to^p Q$, finite, non-singular.
  - (2) $T^{-1/2}X'u \to^d N(0, \sigma^2 Q)$.

- **Theorem** Given this lemma, we can write:

$$\sqrt{T}(\hat{\beta}-\beta) = \underbrace{(T^{-1}X'X)^{-1}}_{\to Q^{-1}} \underbrace{T^{-1/2}X'u}_{\to^d N(0,\sigma^2 Q)} \to^d N(0, \sigma^2 Q^{-1}).$$

So this is our asymptotic distribution, We could also write:

$$\hat{\beta} \approx N(\beta, \frac{1}{T}\sigma^2 Q^{-1}).$$

Further given OLS residuals, $\hat{u}_T = y - X\hat{\beta}$, we have:

$$\hat{\sigma}_T^2 = \frac{1}{T}\hat{u}_T'\hat{u}_T \to^p \sigma^2.$$

So,

$$\begin{aligned} \hat{\beta} &\approx N(\beta, \frac{1}{T}\hat{\sigma}^2(\frac{1}{T}X'X)^{-1}) \\ &\equiv N(\beta, \hat{\sigma}^2(X'X)^{-1}) \end{aligned}$$

- So even though $\hat{\beta}_{OLS}$ is biased, it is still a consistent estimator for β.

16 Lecture 16: October 25, 2005

16.1 Univariate Linear Dynamic Models

Autoregressive Models with iid Errors

- Recall our model:

$$y_t = a_1 y_{t-1} + \cdots + a_p y_{t-p} + z_{t\cdot}\gamma + u_t,$$

or,

$$y_t = x_{t\cdot}\beta + u_t.$$

- We have shown:

$$\sqrt{T}(\hat{\beta}_{OLS} - \beta) \to^d N(0, \sigma^2 Q^{-1}).$$

- Now we move to a special case.

Autoregressive Distributed Lag Model (ADL)

- Consider the model:

$$y_t = m + a_1 y_{t-1} + \cdots + a_p y_{t-p} + c_0 z_t + \cdots + c_q z_{t-q} + u_t,$$

which is an $ADL(p, q)$ model.

- We can write the model using the lag operator as follows. Consider two lag polynomials:

$$A(L) = 1 - a_1 L - \cdots - a_p L^p,$$

$$C(L) = c_0 + c_1 L + \cdots + c_q L^q.$$

So our model becomes:

$$A(L) y_t = m + C(L) z_t + u_t.$$

- Solve for y_t:

$$y_t = A(L)^{-1} m + A(L)^{-1} C(L) z_t + A(L)^{-1} u_t.$$

- Example. Consider the following $ADL(1, 1)$:

$$y_t = m + a_1 y_{t-1} + c_0 x_t + c_1 x_{t-1} + u_t.$$

Thus, $A(L) = 1 - a_1 L$ and:

$$A(L)^{-1} = \frac{1}{1 - a_1 L} = \sum_{i=0}^{\infty} a_1^i L^i.$$

And our model becomes:

$$y_t = m\sum_{i=0}^{\infty} a_1^i L^i + \sum_{i=0}^{\infty} a_1^i L^i[c_0 x_t + c_1 x_{t-1}] + \sum_{i=0}^{\infty} a_1^i L^i u_t.$$

$$y_t = \frac{m}{1-a_1} + c_0 x_t + c_1 x_{t-1} + a_1 c_0 x_{t-1} + a_1 c_1 x_{t-2} + \cdots + \sum_{i=0}^{\infty} a_1^i u_{t-i}.$$

- Staying with the $ADL(1,1)$ for now, consider the following partials:

$$\frac{\partial y_t}{\partial x_t} = c_0, \quad \frac{\partial y_{t+1}}{\partial x_t} = c_1 + a_1 c_0, \quad \frac{\partial y_{t+2}}{\partial x_t} = a_1(c_1 + a_1 c_0).$$

So we can determine the long run effect of a unit change in x on y by two methods.

  - Method 1. Consider the $lim_{s\to\infty}\frac{\partial y_{t+s}}{\partial x_t}$ which is:

$$\text{Long Run Effect } = \sum_{i=0}^{\infty} a_1^i c_0 + \sum_{i=0}^{\infty} a_1^i c_1 = \frac{c_0 + c_1}{1-a_1}.$$

  - Method 2. In the long run, all variables are constant so the model becomes:

$$\overline{y} = m + a_1\overline{y} + c_0\overline{x} + c_1\overline{x}.$$

$$\overline{y} = \frac{m}{1-a_1} + \underbrace{\frac{c_0+c_1}{1-a_1}}_{Long\ Run\ Effect}\overline{x}.$$

In general we have:

$$\text{Long Run Effect } = \frac{C(1)}{A(1)} = \frac{c_0 + \cdots + c_q}{1 - a_1 - \cdots - a_p}.$$

Error Correction Form (ECF)

- Consider rewriting the y and x data as:

$$y_t = y_{t-1} + y_t - y_{t-1} = y_{t-1} + \Delta y_t,$$

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

Which means our model becomes:

$$y_{t-1} + \Delta y_t = m + a_1 y_{t-1} + c_0[x_{t-1} + \Delta x_t] + c_1 x_{t-1} + u_t.$$

$$\Delta y_t = c_0\Delta x_t + -(1-a_1)y_{t-1} + m + (c_0 + c_1)x_{t-1} + u_t.$$

And in even more complicated form:

$$\Delta y_t = c_0 \Delta x_t + -(1-a_1) \underbrace{\left[y_{t-1} - \frac{m}{1-a_1} - \frac{c_0+c_1}{1-a_1} x_{t-1} \right]}_{LRE} + u_t.$$

So note the term in brackets is the Long Run Effect and this is zero in the long run.

- Thus Δy_t is divided between changes in x and deviations from the long run equilibrium. This is our ECF.
- So how do we estimate this model? Could do OLS as shown last lecture which yields consistent (though biased) estimators. To get unbiasedness, we need assumptions on the error terms (so far only iid) which usually puts too many restrictions on the model. Running OLS is more robust.

16.2 Autoregressive Models with Autocorrelated Disturbances

- Consider the model:

$$y_t = a y_{t-1} + u_t, \quad u_t = \rho u_t + \epsilon_t, \quad |a| < 1, |\rho| < 1, \epsilon_t \sim iid(0, \sigma^2).$$

We could rewrite this as:

$$u_t = \frac{1}{1-\rho L}\epsilon_t = \sum_{i=0}^{\infty} \rho^i \epsilon_{t-i}.$$

$$y_t = \frac{1}{1-aL} u_t = \sum_{j=0}^{\infty} a^j u_{t-j}.$$

Thus combining, we have:

$$y_t = \sum_{i=0}^{\infty} c_i \epsilon_{t-i}, \quad c_i = \sum_{k=0}^{i} a^k \rho^{i-k},$$

where c_i is absolutely summable.

- y_t is called an infinite moving average. Since $E[y_t] = 0$ and $cov(y_t, y_{t+h}) = \gamma(h)$, the process is weakly stationary.
- Also,

$$E[y_{t-1}\epsilon_t] = E_{y_{t-1}}[E(y_{t-1}\epsilon_t | y_{t-1})] = E_{y_{t-1}}[y_{t-1} \underbrace{E(\epsilon_t | y_{t-1})}_{0}] = 0.$$

- Asymptotics:

$$plim_{T\to\infty} \frac{1}{T} \sum_{t=2}^{T} y_{t-1}\epsilon_t = 0.$$

$$plim_{T\to\infty}\frac{1}{T}\sum_{t=2}^{T} y_{t-1}^2 = \gamma(0).$$

$$plim_{T\to\infty}\frac{1}{T}\sum_{t=2}^{T} y_t y_{t-1} = plim_{T\to\infty}\frac{1}{T}\sum_{t=2}^{T} y_{t-1}y_{t-2} = \gamma(1).$$

- So how do we estimate the model? We could try OLS:

$$\hat{a}_{OLS} = \frac{\sum_{t=2}^{T} y_t y_{t-1}}{\sum_{t=2}^{T} y_{t-1}^2} = a + \frac{\sum_{t=2}^{T} u_t y_{t-1}}{\sum_{t=2}^{T} y_{t-1}^2}.$$

  However, $E[u_t y_{t-1}] = E[(\rho u_{t-1} + \epsilon_t)y_{t-1}] = \rho E[u_{t-1}y_{t-1}] \neq 0$. So the OLS estimator would be biased AND inconsistent.

- How about running GLS? Consider lagging the model, multiplying by ρ, and then subtracting the result from the original model. This yields:

$$\underbrace{y_t - \rho y_{t-1}}_{y_t^*} = a\underbrace{(y_{t-1} - \rho y_{t-2})}_{y_{t-1}^*} + \epsilon_t.$$

  Since ϵ has nice properties, we could do OLS of y_t^* on y_{t-1}^* which would be consistent. HOWEVER, we don't know ρ so we would have to estimate ρ and then do feasible GLS. In this case, the limiting distribution of the feasible GLS estimator is NOT equal to the limiting distribution of the GLS estimator. It is still consistent but the limiting distribution depends on how your estimate ρ. So this is not attractive.

- One method that does work is the following. Consider again the transformed model:

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

  Or,

$$y_t = \underbrace{(a+\rho)}_{\pi_1} y_{t-1} - \underbrace{a\rho}_{\pi_2} y_{t-2} + \epsilon_t.$$

  So we could do OLS of y_t on the first and second lag of y_t to get consistent estimates of the π's. Then solve for a and ρ. What about the asymptotic distributions? Consider using the delta method.

- **Delta Method** Consider $\delta = \phi(\gamma)$ with:

$$\sqrt{T}(\hat{\gamma} - \gamma) \to^d N(0, \Sigma).$$

  What is the distribution of $\hat{\delta}$? Consider a first order taylor expansion of ϕ around the true parameter value (using the Mean Value Theorem):

$$\hat{\delta} = \phi(\hat{\gamma}) = \underbrace{\phi(\gamma)}_{\delta} + \frac{\partial\phi(\tilde{\gamma})}{\partial\gamma}(\hat{\gamma} - \gamma),$$

where $\tilde{\gamma}$ is an intermediate value. Thus,

$$\sqrt{T}(\hat{\delta}-\delta)=\underbrace{\frac{\partial\phi(\tilde{\gamma})}{\partial\gamma}}_{\to\partial\phi(\gamma)/\partial\gamma}\underbrace{\sqrt{T}(\hat{\gamma}-\gamma)}_{\to N(0,\Sigma)}.$$

So,

$$\sqrt{T}(\hat{\delta}-\delta)\to^{d} N(0,\frac{\partial\phi(\gamma)}{\partial\gamma}\Sigma\frac{\partial\phi(\gamma)}{\partial\gamma}^{\prime}).$$

- So transforming and running OLS will work. It yields consistent estimates. We could also do maximum likelihood or NLLS.

- **ML and NLLS estimation** Consider transforming the first error term:

$$v_1=\sqrt{1-\rho^2}u_1,$$

and assume:

$$(v_1,\epsilon_2,\ldots,\epsilon_T)\sim N(0,\sigma^2 I_T).$$

So,

$$v_1=\sqrt{1-\rho^2}u_1=\sqrt{1-\rho^2}y_1-\sqrt{1-\rho^2}ay_0-\sqrt{1-\rho^2}z_{1\cdot}\gamma,$$
$$\epsilon_t=y_t-(a-\rho)y_{t-1}+\rho a y_{t-2}-z_{t\cdot}\gamma+z_{t-1\cdot}\rho\gamma,\ t=2,\ldots,T.$$

- This gives us a distribution of the y_t's from a change of variables technique (involving a Jacobian). This in turn yields a likelihood function and after simplification, we have the problem of minimizing the sum of squared residuals ... NLLS!

- Note we can maximize the likelihood function in stages (yielding a "concentrated likelihood function"). First with respect to β and then with respect to α given the maximized β. The theory is as follows. Consider maximizing $f(x,y)$ wrt x and y. We have two simultaneous FOCs:

$$f_x=\frac{\partial f(x,y)}{\partial x}=0,\ \ f_y=\frac{\partial f(x,y)}{\partial y}=0.$$

Instead, consider maximizing with respect to y first, yielding $f_y=0\Rightarrow\tilde{y}=g(x)$, and then:

$$Max_x\ \ f(x,g(x)).$$

FOC:

$$0=f_x+\underbrace{f_y}_{=0}g_x=f_x.$$

So we get both f_x and f_y equal to zero. Thus the solution is the same as solving the FOCs simultaneously.

17 Lecture 17: October 27, 2005

17.1 Univariate Linear Dynamic Models

Finalizing Autocorrelated Disturbances

- Recall with autocorrelated disturbances, we can estimate the model by OLS on the transformed model or via ML/NLLS. Both yield consistent though biased coefficients. They differ only by the determinate of the Jacobian of the transformation.

Stationary Time Series

- Consider a stochastic process, $(z_t : t \in T)$ where t is the index variable and T may be either a discrete or continuous space in possibly more than one dimension. If $T = \Re^2$, then z_t is called a "random field."

- **Definition** Autocovariance Function. This is defined as:

$$\gamma_z(t,s) = cov(z_t, z_s) = E[(z_t - E[z_t])(z_s - E[z_s])] \;\forall\, t, s \in T.$$

- **Definition** Weakly Stationary. A process is weakly stationary if:
  - (1) $E[z_t^2] < \infty \;\forall\, t \in T$.
  - (2) $E[z_t] = E[z_s] = \mu$, so we have constant means.
  - (3) The covariance is shift invariant:

  $$cov(z_t, z_s) = \gamma(t-s) \iff cov(z_t, z_{t+h}) = \gamma(h) = \gamma(-h) \iff cov(z_{t+h}, z_{s+h}) = cov(z_t, z_s).$$

  Note this implies $var(z_t) = \gamma(0)$ is constant across time.

  Note that even if a process is not weakly stationary, its first difference, $y_t = z_t - z_{t-1}$, or growth rate, $y_t = ln(z_t) - ln(z_{t-1})$, may be stationary.

- **Definition** Autocorrelation Function. For a weakly stationary process, consider the correlation:

$$\rho_z(h) = corr(z_t, z_{t+h}) = \frac{cov(z_t, z_{t+h})}{\sqrt{var(z_t) * var(z_{t+h})}}.$$

  Or,

$$\rho_z(h) = \frac{\gamma_z(h)}{\gamma_z(0)}.$$

- **Remark** If $z_t = az_{t-1} + \epsilon_t$ with $\epsilon_t \sim iid(0, \sigma_\epsilon^2)$, then if $a > 0$, the autocovariance function looks like G-17.1. For $a < 0$, the autocovariance bounces back and forth from positive to negative as in the graph.

- **Definition** White Noise. A process (ϵ_t) is white noise if $E[\epsilon_t] = 0$ and $E[\epsilon_t \epsilon_s] = \sigma_\epsilon^2$ only when $t = s$ and zero else. Write:

$$\epsilon_t \sim WN(0, \sigma_\epsilon^2).$$

- Consider the process:

$$z_t = az_{t-1} + \epsilon_t, \;\; \epsilon_t \sim iid(0, \sigma_\epsilon^2), \;\; |a| > 1.$$

Then $(1 - aL)z_t = \epsilon_t$ and we can write:

$$z_t = -[a^{-1}L^{-1}(1 - a^{-1}L^{-1})^{-1}]\epsilon_t.$$

Then,

$$z_t = -\sum_{i=0}^{\infty}(1/a)^i \epsilon_{t+i},$$

which IS stationary! So given the first order difference equation, if we consider the FORWARD solution, we get a stationary process. The backward solution is NOT stationary unless $|a| < 1$. If $a = 1$, we have a random walk and no solution will ever be stationary.

- **Definition** Strict Stationarity. A stochastic process is strictly stationary if:

$$\text{The joint distribution of } (z_{t1}, \ldots, z_{tk}) \text{ is the same as}$$

$$\text{the joint distribution of } (z_{t1+h}, \ldots, z_{tk+h}) \; \forall \; k, h \in \mathcal{Z}.$$

So if you pull off any group of random variables and consider their joint distribution, and then shift them all by the same amount (increase or decrease all their indices by the same amount), the joint distribution is unchanged. Thus all single variables must have the same marginal distribution. Clearly strict stationarity implies weak stationarity if the second moments exist.

Causal and Invertible ARMA Processes

- Suppose (y_t) is a stationary stochastic process (with finite second moments). Then,

$$corr(y_t, y_{t+h}) = corr(y_s, y_{s+h}) = \rho_h,$$

which implies a variance covariance matrix for the process as follows:

$$VC(y_t) = \sigma_y^2 \begin{bmatrix} 1 & \rho_1 & \ldots & \rho_{T-1} \\ \rho_1 & 1 & & \rho_{T-2} \\ \vdots & & \ddots & \vdots \\ \rho_{T-1} & \ldots & \rho_1 & 1 \end{bmatrix}.$$

So we only have T parameters instead of $T(T+1)/2$ as usual. This is better but we still need more structure if we want to estimate the model as T parameters is still too many.

- **Definition** A process (y_t) is $ARMA(p,q)$ if (y_t) is stationary and if $\forall\ t$,

$$y_t - a_1 y_{t-1} - \cdots - a_p y_{t-p} = \epsilon_t + b_1 \epsilon_{t-1} + \cdots + b_q \epsilon_{t-q}.$$

  Or in lag notation:

$$A(L)y_t = B(L)\epsilon_t,$$

  with $(\epsilon_t) \sim WN(0, \sigma_\epsilon^2)$. So there are many solutions to the above difference equation, but it is only ARMA if (y_t) is stationary and the difference equation is satisfied.

- Consider an ARMA process, y_t and consider $y_t^* = y_t - \mu$. Then y_t^* is ARMA if:

$$A(L)(y_t - \mu) = B(L)\epsilon_t.$$

  Take expectations:

$$A(L)E[y_t - \mu] = 0,$$

  and since $A(L) \neq 0$, this means $E[y_t] = \mu$, the mean of the process.

- **Remark** Consider $z_t = z$, $E[z_t] = 0$, $E[z_t^2] = \sigma^2$. Clearly this is stationary but note $\rho_z(h) = 1\ \forall\ h$! So the process has infinite memory!

- **Remark** Suppose you have an $ARMA(p,q)$, z_t, with autocovariance function, $\gamma(h)$, such that:

$$\gamma_z(h) \to 0\ as\ h \to \infty.$$

  Thus the process has declining memory. Then for all $k > 0$, $\exists$ an $ARMA(p,q)$, y_t, with autocovariance function, $\gamma_*(h)$, such that:

$$\gamma_*(h) = \gamma_z(h)\ for\ h = 0, \ldots, k.$$

  So the first k autocovariances are the same. I'm not sure why this is useful.

- **Definition** An $ARMA(p,q)$, $A(L)y_t = B(L)\epsilon_t$ is said to be CAUSAL if $\exists$ a sequence of constants, c_i, which are absolutely summable and:

$$y_t = \sum_{i=0}^{\infty} c_i \epsilon_{t-i} \equiv\ MA(\infty)\ process.$$

Review for Midterm

17.2 Lectures Notes

GLS

- Solution to a difference equation: $y_t = c\lambda^t$, and for a differential equation, $y_t = ce^{\lambda t}$.
- GLS estimator:
$$\hat{\beta}_{GLS} = (X'\Omega^{-1}X)^{-1}X'\Omega^{-1}y.$$
- Aitkem Theorem: GLS estimator is BLUE.
- GLS estimator for σ^2:
$$\tilde{\sigma}^2 = \frac{\tilde{u}'\Omega^{-1}\tilde{u}}{T-K}.$$
- GLS asympototics:
$$\sqrt{T}(\hat{\beta}_{GLS} - \beta) \to^d N(0, \sigma^2 Q^{-1}).$$
$$\hat{\beta}_{GLS} \approx N(\beta, \tilde{\sigma}^2(X'\Omega^{-1}X)^{-1}).$$
$$VC(\hat{\beta}_{GLS}) = \frac{\tilde{\sigma}^2(X'\Omega^{-1}X)^{-1}}{T-K}.$$
- Sufficient conditions for $\tilde{\beta}_{GLS} = \tilde{\beta}_{FGLS}$:
  - (1) $plimT^{-1}X'\hat{\Omega}^{-1}X$ finite, nonsingular.
  - (2) $plimT^{-1}X'\hat{\Omega}^{-1}u = 0$.
- Conditions for the feasible and true GLS estimators to have the same asymptotic distributions:
  - (1) $plimT^{-1}X'(\hat{\Omega}^{-1} - \Omega^{-1})X = 0$.
  - (2) $plimT^{-1}X'(\hat{\Omega}^{-1} - \Omega^{-1})u = 0$.
- Feasible GLS asymptotics:
$$\sqrt{T}(\hat{\beta}_{FGLS} - \beta) \to^d N(0, \sigma^2 lim(T^{-1}X'\Omega^{-1}X)^{-1}).$$
$$\hat{\beta}_{FGLS} \approx N(\beta, \hat{\sigma}^2(X'\hat{\Omega}^{-1}X)^{-1}).$$
$$\hat{\sigma}^2 = \frac{\hat{u}'\hat{\Omega}^{-1}\hat{u}}{T-K}.$$
- Additional condition needed for $\hat{\sigma}^2_F \to \sigma^2$ and $\hat{\sigma}^2 \to \sigma^2$:
$$T^{-1}u'(\hat{\Omega}^{-1} - \Omega^{-1})u \to 0.$$

Autocorrelation and Heteroskedasticity

- AR(1) like $u_t = \rho u_{t-1} + \epsilon_t$,
$$u_t = \sum_{i=0}^{\infty} \rho^i \epsilon_{t-i}.$$
$$E[u_t^2] = \sigma_\epsilon^2 \frac{1}{1-\rho^2}.$$
$$E[u_t u_{t+s}] = \sigma_\epsilon^2 \frac{\rho^{|s|}}{1-\rho^2}.$$

- For an AR(1), if ρ is unknown, estimate it with the Cochrane Orcutt Method. Or Hildreth-Lu (grid search), Durbin Method (regress y on lag of y, X and lag of X), or do maximum likelihood.

- Tests for Autocorrelation: Durbin Watson. Run OLS on model and save resids. Calculate:
$$d = \frac{\sum_{t=2}^{T} (\hat{u}_t - \hat{u}_{t-1})^2}{\sum_{t=1}^{T} \hat{u}_t^2}.$$
If $d = 2$, $\rho = 0$, fail to reject. Other tests include the Breusch-Godfrey LM test and the Box Pierce text.

- Under heteroskedasticity, OLs will be unbiased and consistent but inefficient.

- Estimating parameter efficiency without knowing the structure of the variance. White (1980) proposed running OLS, saving resides and then calculating:
$$\tilde{\psi} = T^{-1} \sum_t \tilde{u}_t^2 X_{t\cdot}' X_{t\cdot}.$$
Use this for the term $T^{-1} X' \Sigma X$ in the equation:
$$VC(\hat{\beta}) = (X'X)^{-1} X' \Sigma X (X'X)^{-1},$$
to get an estimate of the variance of our estimator. These are the ROBUST standard errors.

- Tests for Heteroskedasticity. White Test: Run OLS and save resids. Run OLS of squared resids on a functions of the x's. Then,
$$TR^2 \sim \chi^2(p-1),$$
where p is the number of regressors. Other tests include Breusch Pagan LM test, the Goldfeld Quandt test where we split the sample and calculate:
$$GQ = \frac{ESS_2}{ESS_1} \sim F(\frac{T-r}{2} - K, \frac{T-r}{2} - K).$$

Seeming Unrelated Regression

- $E[uu'] = \Omega = \Sigma \otimes I_T$. Estimator:

$$\hat{\beta}_{GLS} = (X'(\Sigma^{-1} \otimes I_T)X)^{-1}X'(\Sigma^{-1} \otimes I_T)y.$$

Panel Data Models

- 2 way error model:

$$y = e_{NT}\alpha + (I_N \otimes e_T)\mu + (e_N \otimes I_T)\lambda + X\beta + v.$$

- Note $J_{NT} = e_{NT}e'_{NT} = (e_T \otimes e_N)(e'_T \otimes e'_N) = e_N e'_N \otimes e_T e'_T$.
- For fixed effects, remember to restrict the sum of the mus and lambdas to be 0 or we have perfect multicollinearity.
- Fixed Effects Asymptotics (least squares dummy variable estimator):

$$\sqrt{NT}(\hat{\beta}_{LSDV} - \beta) \to^d N(0, lim\sigma_v^2((NT)^{-1}X'Q_0X)^{-1}).$$

This estimator is unbiased and consistent!

- Test for fixed effects with an F-test.
- Random effects: include the individual and time specific effects in the error because we assume they are random and DO NOT depend on the x's. If RE model is correct, GLS is most efficient:

$$\hat{\delta}_{GLS} = (Z'\Omega^{-1}Z)^{-1}Z'\Omega^{-1}y.$$

- Random effects asymptotics:

$$\sqrt{NT}(\hat{\beta}_{GLS} - \beta) \to^d N(0, lim\sigma_v^2((NT)^{-1}X'Q_0X)^{-1}).$$

This is the SAME the fixed effects asymptotics.

- Prucha's Theorem: as long as you estimate the variances consistently (using OLS for example), the asymptotic distributions of the FGLS and GLS estimators will be the same.
- Test for random effects using a Breusch Pagan LM test.

Hausman Test

- Consider two estimators:

$$\hat{\theta}_T : \quad plim\ \hat{\theta}_T = \theta, \text{ efficient, under } H_0.$$

$$\tilde{\theta}_T: \quad plim\ \tilde{\theta}_T = \theta, \text{ under } H_0 \text{ and } H_1. \text{ But inefficient under } H_0.$$

- Consider the following Hausman test statistic:

$$(\underbrace{\hat{\theta}_T}_{px1} - \tilde{\theta}_T)'[VC(\hat{\theta}_T - \tilde{\theta}_T)]^{-1}(\hat{\theta}_T - \tilde{\theta}_T) \to \chi^2(p).$$

Munlak Model

- Model:

$$y_{ti} = \alpha + \mu_i + \sum_{k=1}^{K} x_{tik}\beta_k + v_{ti},$$

$$\mu_i = \sum_{k=1}^{K} \overline{x}_{\cdot ik}\pi_i + \epsilon_i.$$

- Estimators:

$$\tilde{\beta}_{GLS} = \hat{\beta}_{LSDV} = (X'\overline{\overline{Q}}_0X)^{-1}X'\overline{\overline{Q}}_0y,$$

$$\tilde{\pi}_{GLS} = \tilde{\beta}_B - \hat{\beta}_{LSDV},$$

$$\tilde{\alpha}_{GLS} = \overline{y} - \overline{x}\tilde{\beta}_B,$$

where $\tilde{\beta}_B = (X'\overline{\overline{Q}}_1X)^{-1}X'\overline{\overline{Q}}_1y$. Thus the GLS estimator in this case is just the FIXED effects estimator.

Hausman Taylor Model

- Consider the following model:

$$y_{ti} = \sum_{k=1}^{K} x_{tik}\beta_k + \sum_{l=1}^{L} z_{il}\gamma_l + u_{ti},$$

$$u_{ti} = \mu_i + v_{ti}.$$

- In matrix notation:

$$y = X\beta + Z\gamma + u, \quad u = (I_N \otimes e_T)\mu + v.$$

- Our variance/covariance matrix can be written:

$$E[uu'] = \Omega = E[uu'|X_1, Z_1] = \sigma_v^2\overline{\overline{Q}}_0 + \sigma_1^2\overline{\overline{Q}}_1.$$

- Hausman Taylor consider the following set of instruments:

$$H = [\overline{\overline{Q}}_0X, \overline{\overline{Q}}_1X_1, Z_1] = [\overline{\overline{Q}}_0X_1, \overline{\overline{Q}}_0X_2, \overline{\overline{Q}}_1X_1, Z_1].$$

- So consider transforming the data as follows:

$$y^* = \sigma_v \Omega^{-1/2} y = [I_{NT} - (1-\theta)\overline{\overline{Q}}_1]y$$

$$X^* = \sigma_v \Omega^{-1/2} X = [I_{NT} - (1-\theta)\overline{\overline{Q}}_1]X$$

$$Z^* = \sigma_v \Omega^{-1/2} Z = [I_{NT} - (1-\theta)\overline{\overline{Q}}_1]Z = \theta Z$$

where $\theta = \dfrac{\sigma_v}{\sigma_1}$

- So we need an estmate of θ. We do this in a few steps.
  - (1) First, transform the data according to:

$$\overline{\overline{Q}}_0 y = \overline{\overline{Q}}_0 X\beta + \overline{\overline{Q}}_0 u,$$

    and run OLS yielding consistent estimates of $\tilde{\beta}$.
  - (2) Second, transform the data according to:

$$\overline{\overline{Q}}_1 y - \overline{\overline{Q}}_1 X\hat{\beta} = \underbrace{\overline{\overline{Q}}_1 Z}_{Z}\gamma + \overline{\overline{Q}}_1 u,$$

    and run an IV regression using X_1 and Z_1 as instruments to get a consistent estimate of $\tilde{\gamma}$.
  - (3) Create residuals:

$$\tilde{u} = y - X\tilde{\beta} - Z\tilde{\gamma}.$$

  - (4) Create AOV estimators:

$$\tilde{\sigma}_1^2 = \frac{\tilde{u}'\overline{\overline{Q}}_1\tilde{u}}{N},$$

$$\tilde{\sigma}_v^2 = \frac{\tilde{u}'\overline{\overline{Q}}_0\tilde{u}}{N(T-1)},$$

$$\tilde{\sigma}_\mu^2 = \frac{\tilde{\sigma}_1^2 - \tilde{\sigma}_v^2}{T}.$$

  - (5) Finally, form our estimate of θ:

$$\tilde{\theta} = \frac{\tilde{\sigma}_v}{\tilde{\sigma}_1}.$$

Univariate Linear Dynamic Models

- Schonfeld CLT. We make the following 4 assumptions:
  - (1) $u_t \sim iid(0, \sigma^2)$, $E[u_t^4] < \infty$.
  - (2) The exogenous Z's are uniformly bounded in absolute value.

- (3) The following holds:

$$lim_{T\to\infty}\frac{1}{T-\tau}\sum_{t=1}^{T-\tau} z'_{t.}z_{(t+\tau)}.$$

is finite for all τ and further for $\tau = 0$, $\frac{1}{T}Z'Z$ converges to something finite and non-singular.

- (4) The roots of the characteristic polynomial:

$$\lambda^p - a_1\lambda^{p-1} - \cdots - a_p = 0,$$

are all less than one in absolute value, ie, $|\lambda_i| < 1 \ \forall\ i$. This ensures we have a stable system.

Lemma By a CLT from Schonfeld (1971), under the assumptions above, we have:

- (1) $T^{-1}X'X \to^p Q$, finite, non-singular.
- (2) $T^{-1/2}X'u \to^d N(0, \sigma^2 Q)$.

- **Theorem** Given this lemma, we can write:

$$\sqrt{T}(\hat{\beta} - \beta) = \underbrace{(T^{-1}X'X)^{-1}}_{\to Q^{-1}}\underbrace{T^{-1/2}X'u}_{\to^d N(0,\sigma^2 Q)} \to^d N(0, \sigma^2 Q^{-1}).$$

So this is our asymptotic distribution, We could also write:

$$\hat{\beta} \approx N(\beta, \frac{1}{T}\sigma^2 Q^{-1}).$$

Further given OLS residuals, $\hat{u}_T = y - X\hat{\beta}$, we have:

$$\hat{\sigma}^2_T = \frac{1}{T}\hat{u}'_T\hat{u}_T \to^p \sigma^2.$$

So,

$$\hat{\beta} \approx N(\beta, \hat{\sigma}^2(X'X)^{-1}).$$

So the estimator is biased but consistent.

Autoregressive Distributed Lag Model (ADL)

- Consider the model:

$$y_t = m + a_1 y_{t-1} + \cdots + a_p y_{t-p} + c_0 z_t + \cdots + c_q z_{t-q} + u_t,$$

which is an $ADL(p, q)$ model.

- Method 2. In the long run, all variables are constant so the model becomes:

$$\overline{y} = m + a_1\overline{y} + c_0\overline{x} + c_1\overline{x}.$$

$$\overline{y} = \frac{m}{1-a_1} + \underbrace{\frac{c_0+c_1}{1-a_1}}_{Long\ Run\ Effect}\overline{x}.$$

- In general we have:

$$\text{Long Run Effect } = \frac{C(1)}{A(1)} = \frac{c_0+\cdots+c_q}{1-a_1-\cdots-a_p}.$$

Error Correction Form (ECF)

- Consider rewriting the y and x data as:

$$y_t = y_{t-1} + y_t - y_{t-1} = y_{t-1} + \Delta y_t,$$

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

- Rearrange:

$$\Delta y_t = c_0\Delta x_t + -(1-a_1)\underbrace{\left[y_{t-1} - \frac{m}{1-a_1} - \frac{c_0+c_1}{1-a_1}x_{t-1}\right]}_{LRE} + u_t.$$

Autoregressive Models with Autocorrelated Disturbances

- Consider the model:

$$y_t = ay_{t-1} + u_t, \quad u_t = \rho u_t + \epsilon_t, \quad |a| < 1, |\rho| < 1, \epsilon_t \sim iid(0,\sigma^2).$$

- We could rewrite this as:

$$u_t = \frac{1}{1-\rho L}\epsilon_t = \sum_{i=0}^{\infty}\rho^i\epsilon_{t-i}.$$

$$y_t = \frac{1}{1-aL}u_t = \sum_{j=0}^{\infty}a^j u_{t-j}.$$

Thus combining, we have:

$$y_t = \sum_{i=0}^{\infty}c_i\epsilon_{t-i}, \quad c_i = \sum_{k=0}^{i}a^k\rho^{i-k},$$

where c_i is absolutely summable.

- How do we estimate? OLS is biased and inconsistent. GLS doesn't have good asymptotics. Cnosider the Delta method.
- **Delta Method** Consider $\delta = \phi(\gamma)$ with:

$$\sqrt{T}(\hat{\gamma} - \gamma) \to^d N(0, \Sigma).$$

$$\hat{\delta} = \phi(\hat{\gamma}) = \underbrace{\phi(\gamma)}_{\delta} + \frac{\partial \phi(\tilde{\gamma})}{\partial \gamma}(\hat{\gamma} - \gamma),$$

$$\sqrt{T}(\hat{\delta} - \delta) = \underbrace{\frac{\partial \phi(\tilde{\gamma})}{\partial \gamma}}_{\to \partial\phi(\gamma)/\partial\gamma} \underbrace{\sqrt{T}(\hat{\gamma} - \gamma)}_{\to N(0,\Sigma)}.$$

So:

$$\sqrt{T}(\hat{\delta} - \delta) \to^d N(0, \frac{\partial \phi(\gamma)}{\partial \gamma} \Sigma \frac{\partial \phi(\gamma)}{\partial \gamma}').$$

Stationary Time Series

- **Definition** Autocovariance Function. This is defined as:

$$\gamma_z(t, s) = cov(z_t, z_s) = E[(z_t - E[z_t])(z_s - E[z_s])] \;\forall\, t, s \in T.$$

- **Definition** Weakly Stationary. A process is weakly stationary if:
  - (1) $E[z_t^2] < \infty \;\forall\, t \in T$.
  - (2) $E[z_t] = E[z_s] = \mu$, so we have constant means.
  - (3) The covariance is shift invariant:

$$cov(z_t, z_s) = \gamma(t-s) \iff cov(z_t, z_{t+h}) = \gamma(h) = \gamma(-h) \iff cov(z_{t+h}, z_{s+h}) = cov(z_t, z_s).$$

  Note this implies $var(z_t) = \gamma(0)$ is constant across time.

- **Definition** Autocorrelation Function. For a weakly stationary process, consider the correlation:

$$\rho_z(h) = corr(z_t, z_{t+h}) = \frac{cov(z_t, z_{t+h})}{\sqrt{var(z_t) * var(z_{t+h})}}.$$

Or,

$$\rho_z(h) = \frac{\gamma_z(h)}{\gamma_z(0)}.$$

- **Definition** Strict Stationarity. A stochastic process is strictly stationary if:

$$\text{The joint distribution of } (z_{t1}, \ldots, z_{tk}) \text{ is the same as}$$

$$\text{the joint distribution of } (z_{t1+h}, \ldots, z_{tk+h}) \;\forall\, k, h \in \mathcal{Z}.$$

Causal and Invertible ARMA Processes

- **Definition** A process (y_t) is $ARMA(p,q)$ if (y_t) is stationary and if $\forall\ t$,

$$y_t - a_1 y_{t-1} - \cdots - a_p y_{t-p} = \epsilon_t + b_1 \epsilon_{t-1} + \cdots + b_q \epsilon_{t-q}.$$

  So there are many solutions to the above difference equation, but it is only ARMA if (y_t) is stationary and the difference equation is satisfied.

- **Remark** Consider $z_t = z$, $E[z_t] = 0$, $E[z_t^2] = \sigma^2$. Clearly this is stationary but note $\rho_z(h) = 1\ \forall\ h$! So the process has infinite memory!

- **Definition** An $ARMA(p,q)$, $A(L)y_t = B(L)\epsilon_t$ is said to be CAUSAL if $\exists$ a sequence of constants, c_i, which are absolutely summable and:

$$y_t = \sum_{i=0}^{\infty} c_i \epsilon_{t-i} \equiv\ MA(\infty)\ process.$$

17.3 Practice Exam Notes

- Only take the expectations operator inside if the terms are absolutely summable! Check this.

- Proposition. If $plim(z_T - w_T) = 0$ and $w_T \to^d w$, then $z_T \to^d w$.

- Lindeberg Feller CLT. Given,

$$w_T = T^{-1/2} \sum_t \frac{x_t u_t}{\sigma_t^2},$$

  with $\dfrac{u_t}{\sigma_t} \sim iidN(0,1)$ and:

$$T^{-1} \sum_t \frac{x_t^2}{\sigma_t^2} \to^p A_1,$$

  then:

$$w_T \to^d N(0, A_1).$$

- Trinity of test procedures. See Handout.

- CLT 4.7. $u_t \sim iid(0,\sigma^2)$ and $limT^{-1}X'X = Q$. Then $T^{-1/2}X'u \to N(0, \sigma^2 Q)$.

- For AR(1) errors, transform $t = 1$ by $\sqrt{1-\rho^2}$.

- If errors are a function of the β's in a SUR, first do OLS to get a consistent estimator for β. Save the resids. Estimate $\hat{\sigma}^2$ and tranform the data with it. Calculate $\hat{\Sigma}$. Create $\hat{\beta}_{SUR}$ which is a GLS estimator.

- Note $Var(X) = E[X^2] - [E[X]]^2$. If $u_t \sim (0, \sigma^2\Omega)$ and $\Omega = diag(\omega_t^2)$, suppose we find to find $plim\frac{u'u}{T} = plim T^{-1}\sum u_t^2 = plim\phi$. Then,

$$E[\phi] = T^{-1}\sum E[u_t^2] = T^{-1}\sum \sigma^2\omega_t^2 = \sigma^2\sum w_t^2/T \to^p \sigma^2.$$

And for the variance,

$$Var(\phi) = \frac{1}{T^2}\sum_{t=1}^{T} Var(u_t^2).$$

$$\vdots$$

18 Lecture 18: November 1, 2005

18.1 Univariate Linear Dynamic Models

More on $ARMA(p,q)$ Models

- Consider a causal $ARMA(p,q)$process,

$$A(L)y_t = B(L)\epsilon_t,$$

with,

$$y_t = C(L)\epsilon_t = \sum_{i=0}^{\infty} c_i \epsilon_{t-i},$$

so we can write the process as an infinite order Moving Average process where the c_i's are absolutely summable.

- Suppose $y_t = \epsilon_t$ is the true process. We could write this as:

$$(1 - aL)y_t = (1 + bL)\epsilon_t,$$

or,

$$y_t = ay_{t-1} + \epsilon_t + b\epsilon_{t-1},$$

but ONLY if $a = -b$. This is necessary to make the new process consistent with the true process. So if $a = -b$, we have a manifold of solutions to the equation and we have an identification problem. So we want to rule out the possibility that $a = -b$. This means we need $A(L)$ and $B(L)$ to have NO COMMON ROOTS!.

- **Theorem** Given an $ARMA(p,q)$,

$$A(L)y_t = B(L)\epsilon_t,$$

where $A(z)$ and $B(z)$ have no common roots. Then y_t is causal, ie,

$$y_t = C(L)\epsilon_t = \sum_{i=0}^{\infty} c_i \epsilon_{t-i},$$

where the c_i's are absolutely summable, IF AND ONLY IF,

$$A(z) \neq 0 \; \forall \; |z| \leq 1.$$

So this means the process is stable. We could also write $A(z) = 0$ implies $z > 1$.

- **Remark** Also, we can write the $C(z)$ polynomial as follows:

$$C(z) = \frac{B(z)}{A(z)}, \; |z| \leq 1.$$

- Example. Consider $A(z) = 1 - az$ and $B(z) = 1 + bz$ with $|z| < 1$. Then our process is:

$$y_t = ay_{t-1} + \epsilon_t + b\epsilon_{t-1}.$$

So,

$$\begin{aligned}
C(z) &= \frac{B(z)}{A(z)} \\
&= \frac{1+bz}{1-az} \\
&= \sum_{i=0}^{\infty} a^i z^i (1+bz) \\
&= 1 + az + a^2z^2 + \cdots + bz + abz^2 + a^2bz^3 + \ldots \\
&= \underbrace{1}_{c_0} + \underbrace{(a+b)}_{c_1} z + \underbrace{a(a+b)}_{c_2} z^2 + \underbrace{a^2(a+b)}_{c_3} z^3 + \ldots
\end{aligned}$$

So we can find the $C(z)$ polynomial one element at a time using this method.

- There is another method of equating coefficients to find the $C(z)$ polynomial. Consider,

$$\begin{aligned}
A(z)C(z) &= B(z) \\
(1-az)(c_0 + c_1z + c_2z^2 + \ldots) &= 1 + bz \\
c_0 + c_1z + c_2z^2 + c_3z^3 + \cdots - azc_0 - ac_1z^2 - ac_2z^3 - \ldots &= 1 + bz \\
c_0 + (c_1 - ac_0)z + (c_2 - ac_1)z^2 + (c_3 - ac_2)z^3 \ldots &= 1 + bz
\end{aligned}$$

So $c_0 = 1$, $c_1 - ac_0 = b$, $c_2 - ac_1 = 0$, etc. For more complicated $A(z)$ or $B(z)$ polynomials, see notes.

- **Definition** Consider an $ARMA(p, q)$,

$$A(L)y_t = B(L)\epsilon_t,$$

with,

$$\epsilon_t = \sum_{j=0}^{\infty} \pi_j y_{t-j},$$

where the π's are absolutely summable. Then (y_t) is invertible.

18.2 Autocorrelation and Partial Autocorrelation Functions of ARMA Processes

- Assume (y_t) is a causal $ARMA(p,q)$, so the autocovariance function is:

$$\gamma(h) = cov(y_t, y_{t+h}) = \sigma_\epsilon^2 \sum_{j=0}^{\infty} c_j c_{j+|h|}.$$

The autocorrelations are defined as:

$$\rho(h) = corr(y_t, y_{t+h}) = \frac{\gamma(h)}{\gamma(0)}.$$

- We now consider the autocorrelations for different MA and AR processes.

Autocorrelations of an $MA(q)$ Process

- Consider an $MA(1)$ process:

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

- Variance:

$$\gamma(0) = \sigma_\epsilon^2(1+b^2).$$

- Autocovariances:

$$\gamma(1) = cov(y_t, y_{t-1}) = E[y_t y_{t-1}] = E[(\epsilon_t + b\epsilon_{t-1})(\epsilon_{t-1} + b\epsilon_{t-2})] = b\sigma_\epsilon^2.$$

$$\gamma(2) = E[y_t y_{t-2}] = E[(\epsilon_t + b\epsilon_{t-1})(\epsilon_{t-2} + b\epsilon_{t-3})] = 0.$$

And in fact $\gamma(h) = 0$ for $h \geq 2$.

- So the autocorrelations cut off after the maximum lag on the $B(z)$ polynomial for an $MA(q)$ process.

Autocorrelations of an $AR(1)$ Process

- Consider an $AR(1)$ process:

$$y_t - ay_{t-1} = \epsilon_t.$$

Or,

$$y_t = \frac{1}{1-aL}\epsilon_t = \sum_{i=0}^{\infty} a^i \epsilon_{t-i} = \sum_{j=0}^{\infty} c_j \epsilon_{t-j},$$

where $c_j = a^i$.

- Autocovariances:

$$\gamma(h) = \sigma_\epsilon^2 \sum_{j=0}^{\infty} c_j c_{j+|h|} = \sigma_\epsilon^2 \sum_{j=0}^{\infty} a^{2j+|h|},$$

or,

$$\gamma(h) = \sigma_\epsilon^2 \left(\frac{a^{|h|}}{1-a^2} \right) = \gamma(0)a^{|h|}.$$

- So the autocorrelations do NOT cut off but they do decline. Also, our autocorrelation function is:

$$\rho(h) = \frac{\gamma(h)}{\gamma(0)} = a^{|h|}.$$

- Alternatively, we can find the autocovariances and autocorrelations another way. We'll use the Yule - Walker Equations. Consider multiplying the process by y_{t-h} for $h \geq 0$,

$$y_t y_{t-h} - a y_{t-1} y_{t-h} = \epsilon_t y_{t-h}.$$

Take expectations:

$$E[y_t y_{t-h}] - aE[y_{t-1} y_{t-h}] = E[\epsilon_t y_{t-h}].$$

Thus we have:

$$\begin{aligned} h = 0: &\quad \gamma(0) - a\gamma(1) = \sigma_\epsilon^2. \\ h = 1: &\quad \gamma(1) - a\gamma(0) = 0. \\ h = 2: &\quad \gamma(2) - a\gamma(1) = 0. \\ &\quad \vdots \\ h = h: &\quad \gamma(h) - a\gamma(h-1) = 0. \end{aligned}$$

And this is a system of difference equations which has a solution,

$$\gamma(h) = ca^h.$$

Note that $\gamma(0) = ca^0 = c$. So,

$$\gamma(h) = \gamma(0)a^h.$$

Noting that $\gamma(0) = \dfrac{\sigma_\epsilon^2}{1-a^2}$, we have found our autocovariance function.

- We can also divide through the difference equations above by $\gamma(0)$ to get the Yule-Walker equations:

$$\begin{aligned} h = 1: &\quad \rho(1) - a = 0. \\ h = 2: &\quad \rho(2) - a\rho(1) = 0. \\ &\quad \vdots \\ h = h: &\quad \rho(h) - a\rho(h-1) = 0. \end{aligned}$$

So again, the solution to the difference equation is:

$$\rho(h) = ca^h,$$

Note that $1 = \rho(0) = c$, so

$$\rho(h) = a^h,$$

which is same autocorrelation function we found above.

Autocorrelations of an $AR(p)$ Process

- Consider an $AR(p)$ process:

$$y_t - a_1 y_{t-1} - \cdots - a_p y_{t-p} = \epsilon_t.$$

Or,

$$A(L)y_t = \epsilon_t.$$

- Thus we can find the $C(z)$ polynomial as follows:

$$C(z) = \frac{B(z)}{A(z)} = \frac{1}{A(z)}.$$

For $A(z) = 1 - a_1 z - a_2 z^2$,

$$\begin{aligned}
1 &= A(z)C(z) \\
&= (1 - a_1 z - a_2 z^2)(c_0 + c_1 z + c_2 z^2 + c_3 z^3) + \ldots) \\
&= c_0 + c_1 z + c_2 z^2 + c_3 z^3 + \cdots - a_1 c_0 z - a_1 c_1 z^2 - a_1 c_2 z^3 - \cdots - a_2 c_0 z^2 - a_2 c_1 z^3 - \ldots \\
&= c_0 + (c_1 - a_1 c_0)z + (c_2 - a_1 c_1 - a_2 c_0)z^2 + (c_3 - a_1 c_2 - a_2 c_1)z^3 + \ldots
\end{aligned}$$

So $c_0 = 1$, $c_1 - a_1 c_0 = 0$, $c_2 - a_1 c_1 - a_2 c_0 = 0$, $c_3 - a_1 c_2 - a_2 c_1 = 0$, etc.

- We can also use the Yule-Walker equations to find the autocovariances and autocorrelations.

- We note that, as is evident from the above results, our findings for the $AR(1)$ case generalize to the $AR(p)$ case: The autocorrelation function of an $AR(p)$ process declines in absolute value to zero as h tends to infinity, but $\rho(h) \neq 0$ even for large values of h. The speed with which $|\rho(h)|$ tends to zero depends on the magnitude of largest eigenroot (in absolute value). Of course, the same statements also hold for the autocovariance function.

19 Lecture 19: November 8, 2005

19.1 Partial Autocorrelation Function

- **Definition** The partial autocorrelation function, say $\alpha(\cdot)$, of a stationary process, (y_t), is defined as:
$$\alpha(1) = corr(y_2, y_1),$$
$$\alpha(h) = corr(y_{h+1} - \tilde{y}_{h+1|1,y_2,\ldots,y_h}, y_1 - \tilde{y}_{1|1,y_2,\ldots,y_h}),\ h \geq 2.$$
Note $\alpha(1)$ is defined for y_2 and y_1 but it equals $corr(y_3, y_2)$, $corr(y_4, y_3)$, etc due to stationarity. So the partial autocorrelation function measures the correlation between the unexplained component using our estimator of the best linear prediction of y.

- The partial autocorrelations (along with the normal autocorrelation functions) help us pin down the lag length (p and q) that we want to use in our model.

- **Theorem** Let (y_t) be a zero mean stationary process with $\gamma(0) > 0$ and $lim_{h\to\infty}\gamma(h) = 0$. Consider the following systems of equations for various choices of h
For $h = 1$:
$$\rho(0)a_{11} = \rho(1).$$
For $h = 2$:
$$\rho(0)a_{12} + \rho(1)a_{22} = \rho(1).$$
$$\rho(1)a_{12} + \rho(0)a_{22} = \rho(2).$$
For $h = 3$:
$$\rho(0)a_{13} + \rho(1)a_{23} + \rho(2)a_{33} = \rho(1).$$
$$\rho(1)a_{13} + \rho(0)a_{23} + \rho(1)a_{33} = \rho(2).$$
$$\rho(2)a_{13} + \rho(1)a_{23} + \rho(0)a_{33} = \rho(3).$$
And so on. Don't ask me where this system comes from, but given these systems (all of them), it must be that $a_{11} = \alpha(1)$, $a_{22} = \alpha(2)$, $a_{33} = \alpha(3)$ and generally,
$$a_{kk} = \alpha(k).$$
So we will use this system of equations to find out partial autocorrelations (the α's).

- Example. Consider an $AR(1)$,
$$y_t - a_1 y_{t-1} = \epsilon_t.$$
Then:
$$y_t y_{t-1} - a_1 y_{t-1} y_{t-1} = \epsilon_t y_{t-1}.$$
Expectations:
$$\rho(1) - a_1\rho(0) = 0.$$
Rearrange:
$$a_1\rho(0) = \rho(1),$$

the Yule-Walker (Y-W) equation. So Y-W MUST hold. Comparing this to the $h = 1$ system above, it must be that,

$$a_{11} = a_1 = \alpha(1) \neq 0.$$

From the $h = 2$ system:

$$a_{12} = a_1, \; and \; a_{22} = 0 = \alpha(2).$$

So the first partial autocorrelation is just a_1 and the second (and higher) are all zero.

- Example. Consider an $AR(2)$,

$$y_t - a_1 y_{t-1} - a_2 y_{t-2} = \epsilon_t.$$

Then:

$$y_t y_{t-1} - a_1 y_{t-1} y_{t-1} - a_2 y_{t-2} y_{t-1} = \epsilon_t y_{t-1}.$$
$$y_t y_{t-2} - a_1 y_{t-1} y_{t-2} - a_2 y_{t-2} y_{t-2} = \epsilon_t y_{t-2}.$$

Expectations:

$$\rho(1) - a_1 \rho(0) - a_2 \rho(1) = 0.$$
$$\rho(2) - a_1 \rho(1) - a_2 \rho(0) = 0.$$

Rearrange:

$$a_1 \rho(0) + a_2 \rho(1) = \rho(1),$$
$$a_1 \rho(1) + a_2 \rho(0) = \rho(2),$$

the Yule-Walker (Y-W) equations. Again compare these with the system above to see that $\alpha(2) = a_2 \neq 0$, $\alpha(1) \neq 0$ though we can't pin it down exactly (yet), and $\alpha(h) = 0$ for $h \geq 3$.

- So for an $AR(p)$ procss, the partial autocorrelations cut off after the maximum lag length. For an $MA(p)$ process, the partial autocorrelations taper off but never completely go to zero.

- Usually, we do NOT know the $\rho(h)$'s so we need to estimate them. Note for the $AR(2)$ above, we couldn't pin down $\alpha(1)$ exactly, but if we knew the ρ's, then:

$$\hat{\alpha}(1) = \hat{a}_{11} = \frac{\hat{\rho}(1)}{\hat{\rho}(0)}.$$

- **Definition** Suppose $y_t = \mu + \sum_{j=-\infty}^{\infty} c_j \epsilon_{t-j}$ with $\epsilon_t \sim iid(0, \sigma_\epsilon^2)$ and the c_j's absolutely summable, then the process (y_t) is called a LINEAR PROCESS. Note causal ARMAs (dynamically stable in the y's) are linear processes.

- So how should we estimate the ρ's? Consider:

$$\overline{y}_T = \frac{1}{T}\sum_{t=1}^{T} y_t,$$

the sample mean. Then:

$$\hat{\gamma}_T(h) = \frac{1}{T}\sum_{t=1}^{T-h}(y_t - \overline{y}_T)(y_{t+h} - \overline{y}_T),$$

the sample covariance. And therefore:

$$\hat{\rho}_T(h) = \frac{\hat{\gamma}_T(h)}{\hat{\gamma}_T(0)},$$

our estimated autocorrelation function.

- $\overline{y}_T$ is $\sqrt{T}$ consistent. See Prucha's notes.

- **Theorem** It is shown that:

$$\sqrt{T}\begin{pmatrix} \hat{\rho}(1) - \rho(1) \\ \vdots \\ \hat{\rho}(h) - \rho(h) \end{pmatrix} \rightarrow^d N(0, W).$$

Where W comes from Bartlett's formula.

- **Remark** Note a special case of a White Noise process. Consider $H_0 : y_t \sim iid(0, \sigma^2)$, ie white noise. Under the null,

$$\rho(0) = 1, \;\; \rho(h) = 0 \; for \; h \neq 0.$$

This means that $W = I$, the identity matrix! So to do this test, consider the Portmanteau or Box Peirce test. Under the null:

$$T\sum_{s=1}^{h} \hat{\rho}^2(s) \sim \chi^2(h),$$

since we have the sum of squared normals.

19.2 Estimation of $ARMA(p, q)$ Processes

- Consider the model:

$$A(L)y_t = B(L)\epsilon_t, \epsilon_t \sim iid(0, \sigma^2_\epsilon).$$

Assume the process is causal, so:

$$y_t = C(L)\epsilon_t = \sum_{j=0}^{\infty} c_j \epsilon_{t-j}, \;\; C(L) = \frac{B(z)}{A(z)}, \; |z| \leq 1.$$

Assume p and q are KNOWN (will estimate them later). Assume $A(z)$ and $B(z)$ do not have common roots.

- So one method of estimating the a's and b's in the $A(L)$ and $B(L)$ polynomial is as follows. Solve the Yule-Walker equations to get estimates of the a's. Then let:

$$y_t^* = A(L)y_t,$$

using our estimates of the a's. So,

$$y_t^* = \epsilon_t - b_1\epsilon_{t-1} - \cdots - b_q\epsilon_{t-q} = B(L)\epsilon_t,$$

So now y_t^* is an $MA(q)$ process. Compute the autocorrelations of $y_*(h)$ and then solve:

$$\rho_{y_*}(h) = \frac{\sum_{j=0}^{q} b_j b_{j+h}}{\sum_{j=0}^{q} b_j^2} \; for \; h = 1, \ldots, q.$$

So we have q equations and q unknowns and we can solve these for the b's. This method is quick and dirty and should really only be used for starting values of another method.

Maximum Likelihood Estimation of an $ARMA(p, q)$

- Suppose $\epsilon_t \sim N(0, \sigma_\epsilon^2)$ which means $(y_t) \sim N(\cdot)$. So,

$$y(T) = (y_1, \ldots, y_T)' \sim N(0, \Gamma_T(a_1, \ldots, a_p, b_1, \ldots, b_q, \sigma_\epsilon^2)).$$

So what is this crazy variance/covariance matrix look like? Consider the following:

$$\Gamma_T = \begin{bmatrix} \gamma(0) & \gamma(1) & \ldots & \gamma(T-1) \\ \gamma(1) & \gamma(0) & & \vdots \\ \vdots & & \ddots & \vdots \\ \gamma(T-1) & \ldots & \ldots & \gamma(0) \end{bmatrix},$$

where $\gamma(h) = \sigma^2 \sum_{i=0}^{\infty} c_i c_{i+|h|}$.

- So this induces a log-likelihood function:

$$L_T(a_1, \ldots, a_p, b_1, \ldots, b_q, \sigma_\epsilon^2) = const - \frac{1}{2} log(|\Gamma_T|) - \frac{1}{2} y(T)'\Gamma_T^{-1} y(T).$$

And we can maximize this beast to get our estimates of the a's and b's.

- **Theorem** Consistency of the maximum likelihood estimators. Let:

$$\hat{a} = (\hat{a}_1, \ldots, \hat{a}_p)',$$

$$\hat{b} = (\hat{b}_1, \ldots, \hat{b}_q)',$$

be our ML estimators for a casual and invertible $ARMA(p,q)$ process,

$$A(L)y_t = B(L)\epsilon_t, \;\; \epsilon_t \sim (0, \sigma_\epsilon^2).$$

Note that we don't need to assume that the errors are normal so we have a stronger result of a Quasi-ML estimator. Then:

$$\sqrt{T}\begin{pmatrix} \hat{a} - a \\ \hat{b} - b \end{pmatrix} \rightarrow^d N(0, V(a,b)),$$

where $V(a,b)$ is a function of pure AR processes in the $A(L)$ and $B(L)$ polynomials. See Prucha's notes.

20 Lecture 20: November 15, 2005

20.1 Prediction with $ARMA(p,q)$ Models

- Consider the model:
$$A(L)y_t = B(L)\epsilon_t, \ \ \epsilon_t \sim WN(0, \sigma_\epsilon^2)$$
where y_t is a causal and invertible process. We can write this $y_t = C(L)\epsilon_t$, or $\epsilon_t = C(L)^{-1}y_t$, with:
$$C(z) = \sum_{i=0}^{\infty} c_i z^i = \frac{B(z)}{A(z)}, \ |z| \leq 1.$$
Assume $A(z)$, $B(z)$, and of course then $C(z)$ are known.

- The information set at time t consists of :
$$\Omega_t = \{y_t, y_{t-1}, y_{t-2}, \dots\}.$$

- We want to predict y_{t+h}, $h \geq 1$. Denote our prediction $y_{t,h}$. The prediction error is thus $y_{t+h} - y_{t,h}$. One criteria might be to choose a predictor to minimize:
$$Min\ E[(y_{t+h} - y_{t,h})^2|\Omega_t].$$

- We restrict ourselves to linear predictors of the form:
$$y_{t,h} = \psi_{h,0}y_t + \psi_{h,1}y_{t-1} + \psi_{h,2}y_{t-2} + \dots.$$

- Rewrite the predictor in terms of the ϵ's (since the process is causal):
$$y_{t,h} = \psi_{h,0}\epsilon_t + \psi_{h,1}\epsilon_{t-1} + \psi_{h,2}\epsilon_{t-2} + \dots = \Psi_h(L)\epsilon_t.$$

- Since $y_{t+h} = c_0\epsilon_{t+h} + c_1\epsilon_{t+h-1} + \dots$, we can write:
$$\begin{aligned} y_{t+h} - y_{t,h} = c_0\epsilon_{t+h} + c_1\epsilon_{t+h-1} + \dots + c_{h-1}\epsilon_{t+1} \ &+ \ c_h\epsilon_t + c_{h+1}\epsilon_{t-1}\dots \\ &- \ \psi_{h,0}\epsilon_t - \psi_{h,1}\epsilon_{t-1} - \dots \end{aligned}$$

  Or,
$$y_{t+h} - y_{t,h} = c_0\epsilon_{t+h} + c_1\epsilon_{t+h-1} + \dots + c_{h-1}\epsilon_{t+1} + \sum_{i=0}^{\infty}(c_{h+i} - \psi_{h,i})\epsilon_{t-i}.$$
  Then,
$$E[(y_{t+h} - y_{t,h})^2] = (c_0^2 + \dots + c_{h-1}^2) + \sum_{i=0}^{\infty}(c_{h+i} - \psi_{h,i})^2\sigma_\epsilon^2.$$
  Which is minimized for $\psi_{h,i} = c_{h+i}$.

- Thus our best (linear) predictor of y_{t+h} is:

$$\begin{aligned} y_{t,h} &= \sum_{i=0}^{\infty} c_{h+i}\epsilon_{t-i} \\ &= (\sum_{i=0}^{\infty} c_{h+i}L^i)\epsilon_t \\ &= (\sum_{i=0}^{\infty} c_{h+i}L^i)C(L)^{-1}y_t \\ &= C(L)^{-1}(\sum_{i=0}^{\infty} c_{h+i}L^i)y_t \end{aligned}$$

- Example. Consider the $ARMA(1,1)$, *** Solve this out before the final ***

$$y_t - 0.5y_{t-1} = \epsilon_t + 0.3\epsilon_{t-1}.$$

Suppose we know $y_t, y_{t-1}, \ldots$ and want to predict y_{t+h}. Then solve:

$$C(L) = \frac{1+0.3L}{1-0.5L},$$

for the c_0, c_1, etc coefficients and plug into the equation above.

Prediction of an $AR(p)$ Process

- Consider the model

$$y_t = a_1 y_{t-1} + \cdots + a_p y_{t-p} + \epsilon_t.$$

Then suppose we want to find $y_{t,h}$. Consider:

$$y_{t+1} = a_1 y_t + \cdots + a_p y_{t-p+1} + \epsilon_{t+1}.$$

Our best guess for ϵ_{t+1} is ZERO! So, clearly,

$$y_{t,h} = a_1 y_t + \cdots + a_p y_{t-p+1}.$$

How about y_{t+2}. Again consider:

$$y_{t+2} = a_1 y_{t+1} + \cdots + a_p y_{t-p+2} + \epsilon_{t+2}.$$

Which we predict with (for similar reasons):

$$y_{t,2} = a_1 y_{t,1} + \cdots + a_p y_{t-p+2}.$$

- This could have also been found using the method above noting that $C(L) = 1/A(L)$ for an AR process.

20.2 Order Selection of $ARMA$ Models - "Finding Lag Lengths"

- Consider an $ARMA(p_y, q_y)$ process where p_y and q_y are the "true" lag lengths. Ie, these are the minimal number of lags where there are no common roots.
- What if we mess up our lag lengths? Three cases:
  - (1) Too few lags: $p < p_y$ and $q < q_y$. The model is then misspecified (omitted variable bias) so the estimators would be biased and inconsistent.
  - (2) Too many lags: $p > p_y$ and $q > q_y$. The model is then not identified (the likelihood function would have a flat section).
  - (3) One side too long and the other just right. The model IS identified but you've overfit the model.
- So how do you know if you have overfit the model (too many lags?) If your estimates are very sensitive to dropping lags, then you might be overfit.
- So how can we estimate p and q? Two methods: (1) Box Jenkins and (2) Likelihood based criteria.

Box Jenkins Methods for Estimating Lag Lengths

- There are three stages: identification, estimator calculation, and diagnostic checking. We use the autocorrelation and partial autocorrelation functions to estimate p and q. Then estimate the model and look at residuals. Run a Portmanteau/Box Pierce test on the residuals to see if they are white noise. If they are, we're golden. This test has little power.

Likelihood based on Information Criteria

- Consider picking a (p, q), estimating the model by ML, and computing the variance of the residuals:
$$\hat{\sigma}^2_{\epsilon,T}(p, q).$$
- So we have to minimize the variance of the residuals but also consider the number of lags we include. Thus we need a criteria that punishes us for adding additional lags. We consider two criteria, the Akaike Information Criteria (AIC) and the BIC:
$$AIC(p, q) = log(\hat{\sigma}^2_{\epsilon,T}(p, q)) + \frac{2}{T}(p + q).$$
$$BIC(p, q) = log(\hat{\sigma}^2_{\epsilon,T}(p, q)) + \frac{log(T)}{T}(p + q).$$

Or in general:

$$\Psi(p,q) = log(\hat{\sigma}^2_{\epsilon,T}(p,q)) + \frac{C(T)}{T}(p+q),$$

where $C(T)$ is some functional form. Note the second term in each expression is the penalty function for adding additional lags.

- See G-20.1 for a depiction of the AIC for an AR process.
- **Theorem** See notes for details of the setup. Consider functions, $C(T)$, such that $T^{-1}C(T) \to 0$. The theorem states:
  - (1) If $C(T)/(2\ log\ log\ T) > 1$ for large T, then $\hat{p}_T = p_y$ and $\hat{q}_T = q_y$ almost surely (strongly consistent).
  - (2) If $C(T) \to \infty$, then $Pr(\hat{p}_T = p_y,\ \ \hat{q}_T = q_y) = 1$ (weakly consistent).
  - (3) If $C(T) < M < \infty$, then,

    $$Pr(\hat{p}_T > p_y,\ \ \hat{q}_T > q_y) \geq const \geq 0,$$

    for large T.

  So by (2), the BIC is weakly consistent and by (3), the AIC is INCONSISTENT. Never the less, the AIC still preforms better than the BIC in small samples so we keep it around.

21 Lecture 21: November 17, 2005

21.1 Nonstationary Time Series

- Consider the following $ARMA(p,q)$ with mean, μ:

$$A(L)(y_t - \mu) = B(L)\epsilon_t.$$

So,

$$y_t = \mu + \frac{B(L)}{A(L)}\epsilon_t.$$

- Suppose $\mu = \alpha + \delta t$,

$$y_t = \alpha + \delta t + \frac{B(L)}{A(L)}\epsilon_t.$$

This process is called **Trend Stationary**. δt is a deterministic trend. If you subtract the trend and intercept, you get a stationary process.

- If we let $u_t = \frac{B(L)}{A(L)}\epsilon_t$, then:

$$A(L)u_t = B(L)\epsilon_t,$$

a stationary ARMA process.

- Now suppose u_t is non-stationary. Suppose:

$$(1 - L)A_*(L)u_t = B(L)\epsilon_t,$$

where the roots of $A_*(z)$ lie outside of the unit circle, so $A(z) = (1 - z)A_*(z) = 0$ has one root on the unit circle, $z = 1$. Rewrite this as:

$$A_*(L)(1 - L)u_t = B(L)\epsilon_t.$$

Or,

$$A_*(L)\Delta u_t = B(L)\epsilon_t,$$

so Δu_t is a stationary process. Rewrite u_t as:

$$u_t = \frac{B(L)}{(1 - L)A_*(L)}\epsilon_t,$$

so our y_t process becomes:

$$y_t = \alpha + \delta t + \frac{B(L)}{(1 - L)A_*(L)}\epsilon_t.$$

Multiply both sides by $(1 - L)$:

$$(1 - L)y_t = \alpha - \alpha + \delta * t - \delta * (t - 1) + \frac{B(L)}{A_*(L)}\epsilon_t.$$

$$\Delta y_t = \delta + \Delta u_t.$$

which is an $ARMA$ with mean, δ. Thus differencing both y_t and u_t, we obtained stationary processes. (y_t) is said to be **Difference Stationary** or a **Unit Root Process** or **Intergrated of Order 1**.

Definition A stochastic process (y_t) that satisfies,

$$(1-L)^d y_t = \delta + \Psi(L)\epsilon_t, \;\; \epsilon_t \sim WN(0, \sigma_\epsilon^2), \;\; \sum_{i=0}^{\infty} |\Psi_i| < \infty,$$

where the roots of $\Psi(z) = 0$ are outside the unit circle. Then $u_t = \Psi(L)\epsilon_t$ is stationary and:

$$(y_t) \sim I(d),$$

ie, y_t is integrated of order d. So by differencing an $I(d)$ process d times, we get a stationary process.

- MAYBE: If $(y_t) \sim I(d)$, with $d \geq 1$, then $var(y_t) \to \infty$. Check this.
- **Definition** A special case of integrated processes. If $(1-L)^d y_t$ is a causal and invertible $ARMA$ process, then the original process, y_t, is an Autoregressive Integrated Moving Average process, $ARIMA(p, d, q)$.
- **Remark** Note when we write $(1-L)^2 y_t$, we mean:

$$\begin{aligned} (1-L)^2 y_t &= (1-L)(1-L)y_t \\ &= (1-L)\Delta y_t \\ &= \Delta y_t - \Delta y_{t-1}, \end{aligned}$$

ie, a difference in differences.

Special Case: Random Walk

- Consider the process:

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

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

Express y_{t+s} and solve backwards:

$$\begin{aligned}
y_{t+s} &= \delta + y_{t+s-1} + \epsilon_{t+s} \\
&= \delta + [\delta + y_{t+s-2} + \epsilon_{t+s-1}] + \epsilon_{t+s} \\
&= 2\delta + y_{t+s-2} + + \sum_{i=0}^{1} \epsilon_{t+s-i} \\
&= 2\delta + \delta + y_{t+s-3} + \epsilon_{t+s-2} + \sum_{i=0}^{1} \epsilon_{t+s-i} \\
&= 3\delta + y_{t+s-3} + \sum_{i=0}^{2} \epsilon_{t+s-i} \\
&\vdots \\
&= s\delta + y_t + \sum_{i=0}^{s-1} \epsilon_{t+s-i}
\end{aligned}$$

- If $\delta = 0$, we have a **Random Walk** and our best guess of tomorrow's value is today's realization.
- If $\delta \neq 0$, we have a **Random Walk with Drift** and our best guess of tomorrow's value is today's realization plus δ.
- As we will show, the forecast error grows with how far the forecast is into the future.

Prediction for a Random Walk and Trend Stationary Process

- Consider two processes:

$$\begin{aligned}
(1)\ y_t &= \delta + y_{t-1} + \epsilon_t, && \text{Difference Stationary, Random Walk} \\
(2)\ y_t &= \alpha + \delta t + \epsilon_t, && \text{Trend Stationary}
\end{aligned}$$

where $\epsilon_t \sim iid(0, \sigma^2)$.

- Thus forecasts for period $t + s$ are:

$$\begin{aligned}
(1)\ y_{t+s|t} &= \delta s + y_t \\
(2)\ y_{t+s|t} &= \alpha + \delta(t + s) = \delta s + (\alpha + \delta t)
\end{aligned}$$

- Forecast Errors:

$$\begin{aligned}(1)\ \ E[(y_{t+s} - y_{t+s|t})^2] &= \sigma_\epsilon^2 s \to \infty \\ (2)\ \ E[(y_{t+s} - y_{t+s|t})^2] &= \sigma_\epsilon^2 < \infty\end{aligned}$$

- Propogation of shocks. Consider a shock in period t to ϵ_t. How do the two series respond. Rewrite the processes:

$$\begin{aligned}(1)\ \ y_t &= \delta s + y_{t-1} + \epsilon_{t+s} + \cdots + \epsilon_t \\ (2)\ \ y_t &= \alpha + \delta(t+s) + \epsilon_{t+s}\end{aligned}$$

  Now consider $\dfrac{\partial y_{t+s}}{\partial \epsilon_t}$. We have:

$$(1)\ \ \frac{\partial y_{t+s}}{\partial \epsilon_t} = 1, \text{ for } s = 0, 1, 2, \ldots$$

$$(2)\ \ \frac{\partial y_{t+s}}{\partial \epsilon_t} = \begin{cases} 1 & \text{for } s = 0 \\ 0 & \text{for } s \geq 1 \end{cases}$$

- So the shock is permanent for a random walk (difference stationary) process, while the shock only effects period t of the trend stationary process.

22 Lecture 22: November 22, 2005

22.1 Testing for Unit Roots

- Consider the process:
$$y_t = \alpha + \delta t + u_t, \;\; u_t = \rho u_{t-1} + \epsilon_t.$$
Lag, times ρ, and subtract:
$$\begin{aligned}
y_t - \rho y_{t-1} &= (1-\rho)\alpha + \delta t - \delta\rho(t-1) + u_t - \rho u_{t-1} \\
y_t &= \rho y_{t-1} + (1-\rho)\alpha + \delta(t - \rho t + \rho) + \epsilon_t \\
&= \underbrace{(1-\rho)\alpha + \delta\rho}_{\alpha^*} + \underbrace{\delta(1-\rho)}_{\delta^*} t + \rho y_{t-1} + \epsilon_t \\
&= \alpha^* + \delta^* t + \rho y_{t-1} + \epsilon_t \\
y_t - y_{t-1} + y_{t-1} &= \alpha^* + \delta^* t + \rho y_{t-1} + \epsilon_t \\
\Delta y_t &= \alpha^* + \delta^* t + \underbrace{(\rho - 1)}_{\gamma} y_{t-1} + \epsilon_t \\
\Delta y_t &= \alpha^* + \delta^* t + \gamma y_{t-1} + \epsilon_t
\end{aligned}$$

- So our null hypothesis of unit root can be written:
$$H_0 : \gamma = 0 \;\; or \;\; \rho = 1.$$
The alternative is a model with a deterministic trend.

- Thus to do a Dicky Fuller test for a unit root, regress Δy_t on an intercept, a trend term, and y_{t-1} and then calculate:
$$DF = \frac{\hat{\gamma}}{\hat{\sigma}_{\hat{\gamma}}}.$$
This is not asympototically normal (at least under the null) so you need to consider the Dicky Fuller tables to find your critical values.

- The above was for an $AR(1)$ process. If you had an $AR(2)$, you should include a Δy_{t-1} term on the RHS. If you had an $AR(3)$, you would need Δy_{t-1} and Δy_{t-2}. And so on. These tests are called "Augmented Dicky Fuller Unit Root Tests." The objective is to include as many lags as needed to obtain iid residuals.

22.2 Seasonal $ARIMA$ Models

- If our data has seasonal effects in it, the mean is NOT constant and therefore it is not stationary. So we need to deseasonalize the data which I'm fairly certain is not a word. Consider:
$$z_t = (1 - L^s)^D (1 - L)^d y_t,$$

where the first operator is the seasonal difference operator and the second is the first differencing operator. For example, if $s = 4$, $D = 1$ and $d = 1$, we would apply this to quarterly data to first, say, subtract fourth quarter from fourth quarter to remove the seasonality and then take a first difference of the result to (hopefully) remove all non-stationarity.

- Write our stationary (deseasoned) process as:

$$\underline{a}(L^s)z_t = \underline{b}(L^s)u_t,$$

where the u_t's themselves are a stationary ARMA process, $A(L)u_t = B(L)\epsilon_t$, with $\epsilon_t \sim WN(0, \sigma_\epsilon^2)$.

- Substituting, we have:

$$A(L)\underline{a}(L^s)z_t = B(L)\underline{b}(L^s)\epsilon_t,$$

which is called an $ARIMA(pdq)x(PDQ)$ process. Plug all of this into stata and you're golden.

- So the method is to transform y_t to get z_t and then test z_t for a unit root using DF. Use time series plots for an initial guess of how much differencing and seasoning you should do. If there is a unit root, consider more differencing. You can also look at the autocorrelation function for traces of seasonality as the ACF will have peaks on the season lengths.

22.3 Multivariate Linear Dynamic Models

Dynamic Linear Simultaneous Equation Models

- In these types of systems, OLS will typically be biased. Consider the model:

$$y_t = y_{t-1}A_1 + \cdots + y_{t-p}A_p + \epsilon_t,$$

where $y_t = [y_{t1} \ldots y_{tG}]$ is a $1xG$ row vector. So we have G equations, T observations.

- Write it more complicated:

$$y_{ti} = \sum_{j=1}^{G} y_{t-1,j}a_{ji1} + \cdots + \sum_{j=1}^{G} y_{t-p,j}a_{jip} + \epsilon_{ti},$$

which is a Vector AutoRegression (VAR) of order p.

- For our purposes, we will consider a slightly broader class of models which has the VAR as a special case (to be dealt with in 722). Consider:

$$y_t = y_tB + y_{t-1}A_1 + \cdots + y_{t-p}A_p + w_tD + \epsilon_t,$$

or,

$$y_{ti} = \underbrace{\sum_{j=1}^{G} y_{tj} b_{ji}}_{\textit{simultaneous system}} + \underbrace{\sum_{j=1}^{G} y_{t-1,j} a_{ji1} + \cdots + \sum_{j=1}^{G} y_{t-p,j} a_{jip}}_{\textit{lags of endog}} + \underbrace{\sum_{k=1}^{M} w_{tk} d_{ki}}_{\textit{exog}} + \epsilon_{ti}.$$

Which we write more compactly as:

$$y_{ti} = \sum_{j=1}^{G} y_{tj} b_{ji} + \sum_{j=1}^{K} z_{tj} c_{ji} + u_{ti}, \quad t = 1 \ldots T, i = 1 \ldots G,$$

where the z's contain all lags of endogenous variables and all exogenous variables.

- We CANNOT use OLS because the disturbance term is correlated with the RHS variables. We'll have to use IV.
- Note we can write the model as:

$$Y = YB + ZC + U,$$

which is the structural form, or we can transform:

$$Y = ZC(I - B)^{-1} + U(I - B)^{-1} = Z\Pi + V,$$

the reduced form of the system.

- So consider a simple example of a simultaneous equation system. Consider the two equation system:

$$\begin{aligned} C_t &= \alpha Y_t + \epsilon_t, \quad \epsilon_t \sim iid(0, \sigma_\epsilon^2) && (1) \\ Y_t &= C_t + A_o && (2) \end{aligned}$$

where α is the Marginal Propensity to Consume (MPC) and A_o is autonomous spending (constant).

- Substituting:

$$Y_t = A_o + \alpha Y_t + \epsilon_t.$$

$$Y_t = \frac{A_o}{1 - \alpha} + \frac{\epsilon_t}{1 - \alpha}.$$

So,

$$Cov(Y_t, \epsilon_t) = \frac{\sigma_\epsilon^2}{1 - \alpha} \neq 0,$$

and,

$$plim \, \frac{1}{T} \sum_{t=1}^{T} Y_t \epsilon_t = \frac{\sigma_\epsilon^2}{1 - \alpha} \neq 0.$$

- Estimating via OLS yields:

$$\hat{\alpha} = \frac{\sum C_t Y_t}{\sum Y_t^2} \to^p \alpha + \frac{1-\alpha}{1 + A_o^2/\sigma_\epsilon^2} \neq \alpha,$$

BIASED! Since $\alpha \in (0,1)$, we would typically overestimate the MPC if we just estimated the first equation above.

- For a supply and demand system, we only observe equilibrium price and quantity so we'll need supply shifters and demand shifts if we want to parameterize supply and demand functions. We could consider supply shifters (like rainfall) and demand shifters (like income) to identify the system.

23 Lecture 23: November 29, 2005

23.1 Simultaneous Dynamic Systems of Equations

- Consider a supply and demand system:

$$\begin{aligned} q_t^d &= a_1 + b_1 p_t + c_1 l_t + u_{t1} \\ q_t^s &= a_2 + b_2 p_t + c_2 r_t + u_{t2} \\ q_t^d &= q_t^s = q_t \end{aligned}$$

  where l_t is labor income and r_t is rainfall. Assume $u_{t\cdot} = [u_{t1}, u_{t2}] \sim (0, \Sigma_{2x2})$. This is our structural form.

- If we solve the system for the endogenous variables, we get:

$$\begin{aligned} q_t &= \pi_1 + \pi_2 l_t + \pi_3 r_t + v_{t1} \\ p_t &= \pi_4 + \pi_5 l_t + \pi_6 r_t + v_{t2} \end{aligned}$$

  where the π's and v's are defined in Prucha's notes, but clearly they are functions of the structural parameters: $a_1, b_1, c_1, a_2, b_2, c_2, \sigma_{11}, \sigma_{22}$, and σ_{12}. Denote $v_{t\cdot} \sim iid(0, \Omega)$. This is our reduced form.

- Now let:

$$y_{t\cdot} = [q_t, p_t],$$

$$z_{t\cdot} = [1, l_t, r_t],$$

$$v_{t\cdot} = [v_{t1}, v_{t2}],$$

$$\Pi = \begin{bmatrix} \pi_1 & \pi_4 \\ \pi_2 & \pi_5 \\ \pi_3 & \pi_6 \end{bmatrix}.$$

  So our model becomes:

$$y_{t\cdot} = z_{t\cdot}\Pi + v_{t\cdot}.$$

- Which we can also write:

$$[q_t, p_t] = [\pi_1 + l_t\pi_2 + r_t\pi_3, \pi_4 + l_t\pi_5 + r_t\pi_6] + [v_{t1}, v_{t2}],$$

  or,

$$\begin{bmatrix} y_{1\cdot} \\ \vdots \\ y_{T\cdot} \end{bmatrix} = \begin{bmatrix} z_{1\cdot}\Pi \\ \vdots \\ z_{T\cdot}\Pi \end{bmatrix} + \begin{bmatrix} v_{1\cdot} \\ \vdots \\ v_{T\cdot} \end{bmatrix},$$

$$Y = Z\Pi + V.$$

- Since V is normal, so is Y. Thus,

$$E[Y] = Z\Pi = \mu,$$

since $E[V] = 0$. So Π is identified if Z is of full rank (3 in our case). This is pretty usual (and weak) assumption to impose so it seems viable.

- Also the variance of Y is the same as the variance of V. So our reduced form parameters are identified! As long as we don't have moving averages in the disturbances, we are going to have identified reduced form parameters. We just need iid or autoregressive errors.

- So the next obvious question is: can we back out the structural parameters from our reduced form estimates? In this case, we have 6 π's and 6 structural parameters so it turns out we can. The structural parameters are exactly (uniquely) identified.

- Consider a slight change to the model where we drop rainfall from the supply system:

$$\begin{aligned} q_t^d &= a_1 + b_1 p_t + c_1 l_t + u_{t1} \\ q_t^s &= a_2 + b_2 p_t u_{t2} \\ q_t^d &= q_t^s = q_t \end{aligned}$$

This yields 4 reduced form Π parameters (identified), but we still have 5 structural parameters so we can not identify all of them! It turns out we can identify the supply equation parameters but not those in the demand equation. See G-23.1. We only have a demand shifter (pins down supply) but we don't have supply shifter (to pin down demand).

- What if we drop rainfall from the supply equation but add another variable (in addition to price and labor income) to the demand equation? Now, demand is still not identified because we don't have supply shifters. The supply equation is now OVER identified since we'll get something like:

$$\hat{b}_2 = \frac{\pi_3}{\pi_6} = \frac{\pi_2}{\pi_5}.$$

Which estimate to use? Maybe a linear combination where we weight each possible estimator by finding the most efficient (smallest variance) estimator.

23.2 General Notation, Definitions and Central Limit Theorems

- Denote a_{ij} the $(i,j)^{th}$ element of A. Denote a^{ij} the $(i,j)^{th}$ element of A^{-1}. If A is mxn, then,

$$vec(A) = \begin{bmatrix} a_{\cdot 1} \\ \vdots \\ a_{\cdot T} \end{bmatrix}_{mnx1}$$

- Kronecker Products. If A is mxn and B is pxq,

$$A \otimes B = \begin{bmatrix} a_{11}B & \dots & a_{1m}B \\ \vdots & & \vdots \\ a_{n1}B & \dots & a_{nm}B \end{bmatrix}$$

$$C_1 = A_1 \otimes B_1, C_2 = A_2 \otimes B_2 \Longrightarrow C_1C_2 = A_1A_2 \otimes B_1B_2.$$

$$C = A \otimes B \Longrightarrow C^{-1} = A^{-1} \otimes B^{-1}.$$

$$vec(AB) = (B' \otimes I)vec(A) = (I \otimes A)vec(B).$$

$$vec(A + B) = vec(A) + vec(B).$$

- Example. Suppose Z is TxK and U is TxG. Then,

$$vec(Z'U) = \begin{bmatrix} Z'u_{\cdot 1} \\ \vdots \\ Z'u_{\cdot G} \end{bmatrix} = \begin{bmatrix} Z' & & 0 \\ & \ddots & \\ 0 & & Z' \end{bmatrix} * \begin{bmatrix} u_{\cdot 1} \\ \vdots \\ u_{\cdot G} \end{bmatrix} = (I \otimes Z') * \begin{bmatrix} u_{\cdot 1} \\ \vdots \\ u_{\cdot G} \end{bmatrix}.$$

So if $u \sim (0, \Sigma \otimes I_T)$, then

$$vec(Z'U) = (I \otimes Z') \begin{bmatrix} u_{\cdot 1} \\ \vdots \\ u_{\cdot G} \end{bmatrix} \sim (0, \underbrace{(I \otimes Z')(\Sigma \otimes I)(I \otimes Z)}_{\Sigma \otimes Z'Z}).$$

- **Definition** Endogenous variables are explained within the model. Exogenous variables are explained outside the model (assumed nonstochastic so uncorrelated with the disturbances). Predetermined variables are either lagged dependent or exogenous variables and are uncorrelated with the contemporaneous disturbances.

- So our i^{th} structural equation becomes:

$$y_{ti} = \sum_{j=1}^{G} y_{tj}b_{ji} + \sum_{j=1}^{K} z_{tj}c_{ji} + u_{ti},\ t = 1 \dots T, i = 1 \dots G,$$

with $b_{ii} = 0$ for all i. The z's are predetermined variables.

- We can also write this as:

$$y_{ti} = y_{t\cdot}b_{\cdot i} + z_{t\cdot}c_{\cdot i} + u_{ti},$$

or,

$$y_{\cdot i} = Yb_{\cdot i} + Zc_{\cdot i} + u_{\cdot i},\ i = 1 \dots G.$$

Expand this to get:

$$[y_{\cdot 1}, \dots, y_{\cdot G}] = [Yb_{\cdot 1}, \dots, Yb_{\cdot G}] + [Zc_{\cdot 1}, \dots, Zc_{\cdot G}] + [u_{\cdot 1}, \dots, u_{\cdot G}],$$

or simply,

$$Y = YB + ZC + U.$$

- So in matrix notation we have our **structural form**:

  $$Y = YB + ZC + U.$$

  And solving for Y:

  $$Y = Z\underbrace{C(I-B)^{-1}}_{\Pi} + \underbrace{U(I-B)^{-1}}_{V},$$
  $$Y = Z\Pi + V,$$

  we have our **reduced form**.

- Clearly, the structural system is NOT identified if every variable appears in every equation. Thus, we'll need a lot of restrictions to identify the system.

24 Lecture 24: December 1, 2005

24.1 More on Simultaneous Dynamic Systems of Equations

- Recall our model from last time:

$$y_{\cdot i} = Yb_{\cdot i} + Zc_{\cdot i} + u_{\cdot i},\ i = 1 \dots G,$$

  where $Yb_{\cdot i} = [y_{\cdot 1}b_{1i} + \cdots + y_{\cdot G}b_{Gi}]$. Some of these will be zero, so we want to rewrite the model just in terms of the variables that actually appear in each equation.

- Denote:

$$Y_i = [y_{\cdot j_1}, \dots, y_{\cdot j_{G_i}}],$$

  which in general has G_i columns of those variables actually in the i^{th} equation. Similarly,

$$Z_i = [z_{\cdot l_1}, \dots, z_{\cdot l_{K_i}}],$$

  which in general has K_i columns of those variables actually in the i^{th} equation.

- Then if β_i and γ_i are those parameters corresponding to the variables actually in the equations, we have:

$$y_{\cdot i} = Y_i\beta_i + Z_i\gamma_i + u_{\cdot i},\ i = 1 \dots G,$$

  where Y_i is TxG_i, β_i is G_ix1, Z_i is TxK_i and γ_i is K_ix1.

- If we denote $e_{\cdot 1} = (1, 0, \dots, 0)'$, ie a column of ones with a 1 in the first position, then $Ye_{\cdot 1} = y_{\cdot 1}$, and in general:

$$y_{\cdot j} = Ye_{\cdot j}.$$

  Thus,

$$Y_i = Y[e_{\cdot j_1}, \dots, e_{\cdot j_{G_i}}] = YL_{i1}.$$

  Similarly,

$$Z_i = ZL_{i2}.$$

  Also, it can be shown:

$$b_{\cdot i} = L_{i1}\beta_i,$$

$$c_{\cdot i} = L_{i2}\gamma_i.$$

  The L matrices are called "Selector Matrices."

- So our model in structural form without the zero restrictions is:

$$[y_{\cdot 1}, \dots, y_{\cdot G}] = Y[b_{\cdot 1}, \dots, b_{\cdot G}] + Z[c_{\cdot 1}, \dots, c_{\cdot G}] + [u_{\cdot 1}, \dots, u_{\cdot G}],$$

$$Y = YB + ZC + U.$$

- Imposing the zero restrictions, we have:

$$y_{\cdot i} = Y_i\beta_i + Z_i\gamma_i + u_{\cdot i}.$$

$$y_{\cdot i} = [Y_i, Z_i] * [\beta_i, \gamma_i]' + u_{\cdot i} = X_i \delta_i + u_{\cdot i}.$$

- We can also stack the model as follows:

$$\begin{bmatrix} y_{\cdot 1} \\ \vdots \\ y_{\cdot G} \end{bmatrix} = \begin{bmatrix} X_1 & & \\ & \ddots & \\ & & X_G \end{bmatrix} \begin{bmatrix} \delta_1 \\ \vdots \\ \delta_G \end{bmatrix} + \begin{bmatrix} u_{\cdot 1} \\ \vdots \\ u_{\cdot G} \end{bmatrix}.$$

Which we can write:

$$y = X\delta + u,$$

where $y = vec(Y)$, $u = vec(U)$, and,

$$vec \begin{pmatrix} B \\ C \end{pmatrix} = \begin{bmatrix} L_1 & & \\ & \ddots & \\ & & L_G \end{bmatrix} \delta.$$

So if you know δ, you can get back to B and C.

- **Theorem** Shonfeld Central Limit Theorem for Dynamic Systems of Equations. Consider the following 5 assumptions:
  - (S1) Vectors $u'_{t\cdot} \sim iid(0, \Sigma)$, with Σ nonsingular.
  - (S2) $(I - B)$ is non-singular.
  - (S3) Let W be the submatrix of Z correponding to the (nonstochastic) purely exogenous variables (ie, not the lagged endogenous ones). Then:

$$\underline{Q}^0 = lim_{T\to\infty} T^{-1} W'W \text{ is finite and nonsingular.}$$

$$\underline{Q}^k = lim_{T\to\infty} (T-k)^{-1} \sum_{t=1}^{T-k} w'_{t\cdot} w_{t+k\cdot} \text{ finite for } k \geq 1.$$

  - (S4) Write the model as:

$$Y(I - B) = ZC + U = WC_0 + Y_{-1}C_1 + \cdots + Y_{-H}C_H + U,$$

    where Y_{-i} is the matrix of observations on Y lagged i periods and H is the maximum lag length. We assume the roots of the determinantal equation:

$$det\{\lambda^H (I - B) - \lambda^{H-1} C_1 - \cdots - C_H\} = 0,$$

    are less than 1 in absolute value. That is, the system is STABLE.
  - (S5) The elements of W are uniformly bounded in absolute value.

Then,

$$\underline{Q} = plim_{T\to\infty} T^{-1} Z'Z \text{ is finite and nonsingular,}$$

and,

$$\xi = vec(T^{-1/2}Z'U) = (I \otimes T^{-1/2}Z')u \rightarrow^d N(0, \Sigma \otimes \underline{Q}).$$

- **Remark** Note Σ permits for the disturbances to be contemporaneously correlated over equations as in SUR.

- **Remark** If $Z = W$, nonstochastic, then $E[\xi] = 0$, and

$$E[\xi\xi'] = \Sigma \otimes T^{-1}Z'Z.$$

- **Corollary 1** The theorem also implies:

$$T^{-1/2}Z'u_{\cdot i} \rightarrow^d N(0, \sigma_{ii}\underline{Q}).$$

- **Corollary 2** The theorem also implies:

$$plim\ T^{-1}Z'u_{\cdot i} \rightarrow^p 0.$$

$$plim\ T^{-1}Z'U \rightarrow^p 0.$$

- **Corollary 3** The theorem also implies:

$$plim\ T^{-1}Z'v_{\cdot i} \rightarrow^p 0.$$

$$plim\ T^{-1}Z'V \rightarrow^p 0.$$

- So the predetermined variables can be used as INSTRUMENTS !

24.2 Identification

- Consider the model $Y(I - B) = ZC + U$. If the Z's are nonstochastic, then if U is normal, Y is also normal. Since the Z's contain lagged Y's, we need to do more to get to Y's distribution. We use the tranformation technique and include the Jacobian of the transformation as usual. Even with lagged endogenous variables, since everything is linear, if U is normal, so is Y.

- Write the reduced form of the model

$$Y = Z\Pi + V.$$

  Thus,

$$E[Y] = Z\Pi.$$

  And,

$$var[Y] = var[V] = var[U(I - B)^{-1}] = (I - B')^{-1}\Sigma(I - B)^{-1} = \Omega.$$

- So recall as long as Z is of full rank and we don't have moving average disturbances, the reduced form parameters are always identified. We now consider conditions in which the structural parameters will also be identified. If one equation has a lot of shifters, it is usually the case that these shifters will identify the OTHER equation but the equation with a lot of variables will remain unidentified.

25 Lecture 25: December 6, 2005

25.1 More on Identification

- Recall our relationships from last time:

$$\Pi = C(I - B)^{-1}, \quad \Omega = (I - B')^{-1}\Sigma(I - B)^{-1}.$$

Which we can write:

$$\Pi(I - B) = C.$$

$$(I - B')\Omega(I - B) = \Sigma.$$

So if we know Π and we can get back to B and C, we can then solve for Σ using Ω (which is known). So we will impose some zero restrictions and see if we can solve the first equation for B and C.

- Consider the first equation in an $M = G$ system of equations:

$$y_{\cdot 1} = Y_1\beta_1 + Z_1\gamma_1 + u_{\cdot 1},$$

where $Y = [y_{\cdot 1}, Y_1, Y_1^*]$ and $Z = [Z_1, Z_1^*]$, where starred matrices represent variables in the system but not in equation 1. Thus:

$$Y_1 = [y_{\cdot 2}, \ldots, y_{\cdot M_1+1}], \quad Z_1 = [z_{\cdot 1}, \ldots, z_{\cdot K_1}],$$

so we have M_1 endogenous RHS variables and K_1 exogenous (predetermined) RHS variables.

- Write the relationship above for the first equation:

$$\Pi(i_{\cdot 1} - b_{\cdot 1}) = c_{\cdot 1},$$

Or, if we partition things, we can rewrite this equation as follows:

$$\Pi(i_{\cdot 1} - (1, -\beta_1, 0)') = \Pi \begin{bmatrix} 1 \\ -\beta_1 \\ 0 \end{bmatrix} = \begin{bmatrix} \gamma_1 \\ 0 \end{bmatrix}.$$

So here we are explicitly imposing the zero restrictions. Now rewrite again but partition Π:

$$\begin{bmatrix} \pi_{11} & \pi_{12} & \pi_{13} \\ \pi_{21} & \pi_{22} & \pi_{23} \end{bmatrix} \begin{bmatrix} 1 \\ -\beta_1 \\ 0 \end{bmatrix} = \begin{bmatrix} \gamma_1 \\ 0 \end{bmatrix}.$$

- This yields two equations:

$$\pi_{11} - \pi_{12}\beta_1 = \gamma_1,$$

$$\pi_{21} - \pi_{22}\beta_1 = 0.$$

The first equation says that we can solve for γ_1 if we can solve for β_1. So the second equation is key to identifying β_1. Rewrite:

$$\underbrace{\pi_{22}}_{K-K_1 x M_1} \quad \beta_1 = \pi_{21}.$$

When does this have a unique solution?

- **Rank Condition** A necessary and sufficient condition for identifying β_1, which in turn gives us γ_1, is for π_{22} to have FULL RANK! That is, we need:

$$rank(\pi_{22}) = M_1.$$

  This is the rank condition.

- **Order Condition** A necessary condition for identifying β_1, which in turn gives us γ_1, is for $K - K_1 \geq M_1$, or the number of excluded predetermined variables must exceed the number of included endogenous variables. We can also write this as:

$$M_1 + K_1 \leq K,$$

  or the number of explanatory RHS variables must be less than the total number of predetermined variables. This is the order condition. If $K = M_1 + K_1$, then we say that equation 1 is exactly identified.

- **Remark** Note the reduced form model is:

$$Y = Z\Pi + V.$$

  Or,

$$[y_{\cdot 1}, Y_1, Y_1^*] = Z\Pi + V.$$

  Which implies:

$$Y_1 = Z_1\pi_{12} + Z_1^*\pi_{22} + V_1.$$

  So the π_{22} matrix are the parameters in the reduced form for Y_1 that correspond to the excluded variables.

25.2 OLS Estimation of Structural & Reduced Form Parameters

OLS Estimation of Structural Parameters

- Consider the following lemmatta:

- **Lemma**

$$T^{-1}U'U = T^{-1}\sum_t u'_{t\cdot}u_{t\cdot} \to^p \Sigma,$$

  since the u's are iid, by Khinchine, this quantity converges to its expected value.

- **Lemma**
$$T^{-1}V'V \to^p (I-B')^{-1}\Sigma(I-B)^{-1} = \Omega,$$
since $V = U(I-B)^{-1}$ and applying Khinchine.

- **Lemma**
$$T^{-1}Z'U \to^p 0,$$
by Schonfeld CLT.

- **Lemma**
$$T^{-1}Z'Z \to^p \underline{Q},$$
by assumption.

- **Lemma**
$$T^{-1}Z'V = \underbrace{T^{-1}Z'U}_{\to 0}(I-B)^{-1} \to^p 0.$$

- **Lemma**
$$T^{-1}Y'U = T^{-1}(Z\Pi + V)'U = T^{-1}\Pi'Z'U + T^{-1}V'U = \underbrace{T^{-1}\Pi'Z'U}_{\to 0} + T^{-1}(I-B')^{-1}U'U$$
$$T^{-1}Y'U \to^p (I-B')^{-1}\Sigma \neq 0.$$
Thus OLS is NOT consistent!!

- Now we'll show that OLS is also biased for the structural parameters. Let $X_i = [Y_i, Z_i] = [Y, Z]L_i$, where L_i is a selector matrix. The i^{th} equation is thus:
$$y_{\cdot i} = Y_i\beta_i + Z_i\gamma_i + u_{\cdot i} = X_i\delta_i + u_{\cdot i}.$$
So our OLS estimator is:
$$\hat{\delta}_{i,OLS} = (X_i'X_i)^{-1}X_i'y_{\cdot i} = \delta_i + (X_i'X_i)^{-1}X_i'u_{\cdot i}.$$
Consider some terms:
$$\frac{1}{T}X_i'X_i = \frac{1}{T}L_i'(Y', Z')'(Y, Z)L_i = L_i' \begin{bmatrix} T^{-1}Y'Y & T^{-1}Y'Z \\ T^{-1}Z'Y & T^{-1}Z'Z \end{bmatrix} L_i.$$
So we need the plim of those elements in the matrix. Consider:
$$T^{-1}Y'Y = T^{-1}(Z\Pi+V)'(Z\Pi+V) = T^{-1}(\Pi'Z'Z\Pi+V'Z\Pi+\Pi'Z'V+V'V) \to^p \Pi'\underline{Q}\Pi+\Omega.$$
$$T^{-1}Y'Z = T^{-1}(Z\Pi+V)'Z = T^{-1}(\Pi'Z'Z+V'Z) \to^p \Pi'\underline{Q}.$$
$$T^{-1}Z'Z = \to^p \underline{Q}.$$
So,
$$\frac{1}{T}X_i'X_i \to^p L_i' \begin{bmatrix} \Pi'\underline{Q}\Pi+\Omega & \Pi'\underline{Q} \\ \underline{Q}\Pi & \underline{Q} \end{bmatrix} L_i.$$

Similiarly,

$$\frac{1}{T}X_i'u_{\cdot i} = \frac{1}{T}L_i'(Y', Z')'u_{\cdot i} = L_i'\begin{bmatrix} T^{-1}Y'u_{\cdot i} \\ T^{-1}Z'u_{\cdot i} \end{bmatrix}.$$

So,

$$\frac{1}{T}X_i'u_{\cdot i} \to^p L_i'\begin{bmatrix} (I - B')^{-1}\sigma_{\cdot i} \\ 0 \end{bmatrix}.$$

Finally,

$$\hat{\delta}_{i,OLS} = \delta_i + (X_i'X_i)^{-1}X_i'u_{\cdot i} \neq \delta_i.$$

So, in general, this biases ALL of the δ's, not just those for the endogenous variables!

- So OLS for the structural parameters is biased and inconsistent.

OLS Estimation of Reduced Form Parameters

- Recall our reduced form model:

$$Y = Z\Pi + V.$$

So,

$$\hat{\Pi}_{OLS} = (Z'Z)^{-1}Z'Y = \Pi + (Z'Z)^{-1}Z'V.$$

- Asympototics:

$$\hat{\Pi}_{OLS} = \Pi + \underbrace{(T^{-1}Z'Z)^{-1}}_{\to \underline{Q}}\underbrace{T^{-1}Z'V}_{\to 0} \to^p \Pi.$$

So the reduced form parameters can be estimated consistently by OLS.

- One might think that while consistent, is OLS generating the most efficient estimator? How about doing GLS or SUR on the system? Since the set of explanatory variables is the same in ALL equations, SUR/GLS reduces to OLS so we get the best estimator by just running OLS.
- Not Covered: Indirect Least Squares. See Prucha notes. Estimate the reduced form via OLS and then use your estimates to get consistent estimates of the structural parameters. If an equation is overidentified, you might get multiple estimates of the structural parameters.

26 Lecture 26: December 8, 2005

26.1 Full Information Maximum Likelihood Estimator

- Recall our model from last time:

$$y_{\cdot i} = X_i\delta_i + u_{\cdot i}, \quad X_i = [Y_i, Z_i].$$

  Stack the model yields:

$$y = X\delta + u,$$

  but this model clearly suffers from correlation between the X's and u's. Hence we need to do Instrumental Variables estimation.

- Consider the reduced form model:

$$Y = Z\Pi + V, \quad Y_i = YL_{i1},$$

$$Y_i = Z\Pi L_{i1} + VL_{i1} = Z\Pi_i + V_i.$$

  The best instrument for Y is $E[Y] = Z\Pi$. Thus, for equation i, use $\hat{Y}_i = Z\hat{\Pi}_i$, where $\hat{\Pi}_i$ is some consistent reduced form estimate. Note Z includes all predetermined variables.

- Thus our set of instruments is $\hat{X} = diag_m(\hat{X}_i)$, where $\hat{X}_i = [\hat{Y}_i, Z_i]$, $\hat{Y}_i = \hat{Y}L_{i1}$, and $\hat{Y} = Z\hat{\Pi}$.

- Thus the **Limited Information** Estimator is:

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

  This would be consistent but NOT efficient since $var(u) = \Sigma \otimes I_T$.

- The **Full Information** (IV) estimator is thus as follows:

$$\tilde{\delta} = (\hat{X}'(\hat{\Sigma}^{-1} \otimes I_T)X)^{-1}\hat{X}'(\hat{\Sigma}^{-1} \otimes I_T)y.$$

- And the **Full Information Maximum Likelihood** estimators, $\hat{\delta}$ and $\hat{\Sigma}$, corresponding to normally distributed disturbances, solve the following normal equations:

$$\hat{\delta} = (\hat{X}'(\hat{\Sigma}^{-1} \otimes I_T)X)^{-1}\hat{X}'(\hat{\Sigma}^{-1} \otimes I_T)y \; := \; G(\hat{\Pi}, \hat{\Sigma}^{-1}).$$

$$\hat{X} = diag_m(\hat{X}_i), \; \hat{X}_i = [\hat{Y}_i, Z_i], \; \hat{Y}_i = \hat{Y}L_{i1}, \; \hat{Y} = Z\hat{\Pi}.$$

$$\hat{\Sigma} = T^{-1}\hat{U}'\hat{U}, \; \hat{U} = [Y - Y\hat{B} - Z\hat{C}], \; \hat{\Pi} = \hat{C}(I - \hat{B})^{-1}, vec(\hat{B}, \hat{C})' = L\hat{\delta}.$$

- **Remark** The FIML estimator is asympototically efficient for normal distributions.

- **Theorem 1** For a single equation, if $\tilde{\Pi}_i \to^p \Pi_i$, consistent, then:

$$T^{1/2}(\tilde{\delta}_i - \delta_i) \to^d N(0, \tilde{\sigma}_{ii}(T^{-1}\tilde{X}_i'\tilde{X}_i)^{-1}),$$

where,

$$\tilde{\sigma}_{ii} = T^{-1}(y_{\cdot i} - X_i\tilde{\delta}_i)'(y_{\cdot i} - X_i\tilde{\delta}_i).$$

So whenever $\tilde{\Pi}$ is consistent then $\tilde{\delta}$ will have the same asympototic distribution. Thus,

$$\tilde{\delta}_i \approx N(\delta_i, \tilde{\sigma}_{ii}(\tilde{X}_i'\tilde{X}_i)^{-1}).$$

- **Theorem 2** For a system, if $\tilde{\Pi} \rightarrow^p \Pi$ and $\tilde{\Sigma} \rightarrow^p \Sigma$, consistent, then:

$$T^{1/2}(\tilde{\tilde{\delta}} - \delta) \rightarrow^d N(0, [plim\, T^{-1}\tilde{X}'(\tilde{\Sigma}^{-1} \otimes I_T)\tilde{X}]^{-1}).$$

Thus,

$$\tilde{\tilde{\delta}} \approx N(\delta, (\tilde{X}'(\tilde{\Sigma}^{-1} \otimes I_T)\tilde{X})^{-1}).$$

- **Proof of Theorem 1**

Denote:

$$\tilde{Y} = Z\tilde{\Pi}, \;\; \tilde{Y}_i = Z\tilde{\Pi}_i = Z\tilde{\Pi}L_i.$$

$$X_i = [Y_i, Z_i] = [Y, Z]L_i.$$

$$\tilde{X}_i = [\tilde{Y}_i, Z_i] = [\tilde{Y}, Z]L_i = Z[\tilde{\Pi}, I_K]L_i = Z[\tilde{\Pi}_i, L_{i2}].$$

$$\tilde{R}_i = [\tilde{\Pi}_i, L_{i2}] = [\tilde{\Pi}, I_K]L_i.$$

$$R_i = [\Pi_i, L_{i2}] = [\Pi, I_K]L_i.$$

Consider,

$$\sqrt{T}(\tilde{\delta}_i - \delta) = (T^{-1}\tilde{X}_i'X_i)^{-1}T^{-1/2}\tilde{X}_i'u_{\cdot i}.$$

Consider the first term on the RHS:

$$\begin{aligned}
T^{-1}\tilde{X}_i'X_i &= T^{-1}\tilde{R}_i'Z'[Y, Z]L_i \\
&= \tilde{R}_i'[T^{-1}Z'Y, T^{-1}Z'Z]L_i \\
&= \tilde{R}_i'[T^{-1}Z'(Z\Pi + V), T^{-1}Z'Z]L_i \\
&\rightarrow R_i'[\underline{Q}\Pi, \underline{Q}]L_i \\
&= R_i'\underline{Q}[\Pi, I_K]L_i \\
&= R_i'\underline{Q}R_i
\end{aligned}$$

And,

$$\begin{aligned}
T^{-1/2}\tilde{X}_i'u_{\cdot i} &= T^{-1/2}\tilde{R}_i'Z'u_{\cdot i} \\
&= \tilde{R}_i'T^{-1/2}Z'u_{\cdot i} \\
&\rightarrow R_i' * N(0, \sigma_{ii}\underline{Q}) \\
&= N(0, \sigma_{ii}R_i'\underline{Q}R_i)
\end{aligned}$$

Thus,

$$T^{1/2}(\tilde{\delta}_i - \delta_i) \to^d N(0, \tilde{\sigma}_{ii}(R_i'\underline{Q}R_i)^{-1}).$$

QED.

- Note from our limiting distribution, we also automatically get consistency of our estmator as a by-product. The proof for theorem 2 is analogous.

27 Lecture 27: December 13, 2005

27.1 Limited Information Single Equation Estimation

Two-Stage Least Squares - 2SLS

- Consider the following estimator:

$$\tilde{\delta}_{i,2SLS} = (\tilde{X}_i' X_i)^{-1} \tilde{X}_i y_{\cdot i},$$

where $\tilde{X}_i = (\tilde{Y}_i, Z_i)$, (our instruments), and $\tilde{Y}_i = Z\tilde{\Pi}_i$, where $\tilde{\Pi}_i$ could be estimated as:

$$\tilde{\Pi}_i = \tilde{\Pi}_{i,OLS} = (Z'Z)^{-1} Z'Y_i.$$

Since $Y_i = Z\Pi_i + V_i$, $\tilde{Y}_i = Z\tilde{\Pi}_i = Z(Z'Z)^{-1}Z'Y_i$.

- So this is one way to form our 2SLS estimator. We can also form it several other ways.

- **Proposition** $\tilde{X}_i'\tilde{X}_j = \tilde{X}_i' X_j$.
Proof. Write as:

$$\begin{bmatrix} \tilde{Y}_i'\tilde{Y}_j & \tilde{Y}_i' Z_j \\ Z_i'\tilde{Y}_j & Z_i' Z_j \end{bmatrix} = \begin{bmatrix} \tilde{Y}_i' Y_j & \tilde{Y}_i' Z_j \\ Z_i' Y_j & Z_i' Z_j \end{bmatrix}.$$

So we need to prove $\tilde{Y}_i'\tilde{Y}_j = \tilde{Y}_i' Y_j$ and $Z_i'\tilde{Y}_j = Z_i' Y_j$. Consider the first equality:

$$\begin{aligned} \tilde{Y}_i'\tilde{Y}_j &= L_{i1}'\tilde{Y}'\tilde{Y} L_{j1} \\ &= L_{i1}'(Z(Z'Z)^{-1}Z'Y)'(Z(Z'Z)^{-1}Z'Y)L_{j1} \\ &= L_{i1}'Y'Z(Z'Z)^{-1}Z'YL_{j1} \\ &= L_{i1}'\tilde{Y}'YL_{j1} \\ &= \tilde{Y}_i' Y_j \end{aligned}$$

Also,

$$\begin{aligned} Z_i'\tilde{Y}_j &= L_{i2}'Z'(Z(Z'Z)^{-1}Z'YL_{j1}) \\ &= L_{i2}'Z'YL_{j1} \\ &= Z_i' Y_j \end{aligned}$$

QED.

- The implication of the last proposition is that we can write our estimator as:

$$\tilde{\delta}_{i,2SLS} = (\tilde{X}_i'\tilde{X}_i)^{-1} \tilde{X}_i y_{\cdot i},$$

which is like the estimator from a simple OLS regression of $y_{\cdot i}$ on $\tilde{X}_i$. This motivates why it's called the 2-stage estimator. First regress Y_i on the Z's to get an estimate of

Π_i and then replace the fitted values of $\tilde{Y}_i$ in our model to estimate $\delta_{i,2SLS}$. But this is sort of a strange way to motivate it because statistical packages are really doing the first interpretation above. Both, however, give us the same estimator.

- Also recall our asymptotic properties:

$$\sqrt{T}(\tilde{\delta}_i - \delta_i) \rightarrow^d N(0, \sigma_{ii}^2 [lim\ T^{-1} \tilde{X}_i' \tilde{X}_i]^{-1}),$$

so,

$$\tilde{\delta}_i \sim N(\delta_i, \hat{\sigma}_{ii}(\tilde{X}_i' \tilde{X}_i)^{-1}),$$

with,

$$\hat{\sigma}_{ii} = T^{-1}[y_{\cdot i} - X_i \tilde{\delta}_i]'[y_{\cdot i} - X_i \tilde{\delta}_i].$$

Note in this final estimator, we use the TRUE values of X, not the fitted values.

- Yet another interpretation of 2SLS is as a special case of a GLS estimator. Premultiplying the model by Z' breaks the correlation with the regressors yielding:

$$Z' y_{\cdot i} = Z' X_i \delta_i + Z' u_{\cdot i}.$$

So, here we would have our GLS estimator:

$$\tilde{\delta}_{GLS} = (\underbrace{X_i' Z (Z'Z)^{-1} Z'}_{\tilde{X}_i'} X_i)^{-1} X_i' Z (Z'Z)^{-1} Z' y_{\cdot i}.$$

Again this is just our 2SLS estimator.

- Another interpretation is as a special case of a GMM estimator. Consider minimizing:

$$Min\ Q = (y_{\cdot i} - X_i \delta_i)' Z (Z'Z)^{-1} Z (y_{\cdot i} - X_i \delta_i).$$

This would ALSO yield the 2SLS estimator.

- Finally, if the model was not linear, we could do NLLS using the same objective function but a different functional form for the residuals.

27.2 Full Information System of Equations Estimation

Three-Stage Least Squares - 3SLS

- Our estimator is of the form:

$$\tilde{\delta}_{3SLS} = (\tilde{X}_{OLS}' (\tilde{\Sigma}_{2SLS}^{-1} \otimes I_T) X)^{-1} \tilde{X}_{OLS}' (\tilde{\Sigma}_{2SLS}^{-1} \otimes I_T) y,$$

where $\tilde{X}$ is a block diagonal matrix of all the X's for the G-equation system. The $\tilde{X}_{OLS}$ includes the $\tilde{Y}$ variables generated from running OLS on the reduced form model.

- What about $\tilde{\Sigma}^{-1}_{2SLS}$? A typical element would be:

$$\tilde{\sigma}_{ij,2SLS} = \frac{1}{T}(y_{\cdot i} - X_i \delta_{i,2SLS})'(y_{\cdot i} - X_i \delta_{i,2SLS}).$$

- So again, in stage 1, we regress y on the Z's to estimate Π. In stage 2, we formulate our 2SLS estimator. And in stage 3, we estimate $\tilde{\Sigma}$ and form our 3SLS estimator as above.
- Again, we could also form the 3SLS estimator using GLS or GMM.
- Done.