POOLING TIME SERIES AND CROSS-SECTIONAL DATA

The Model: $y_{it} = \beta_1 x_{it1} + \beta_2 x_{it2} + \cdots + \beta_K x_{itK} + \epsilon_{it}$

$i = 1, \ldots, N$ Cross-Sections
$t = 1, \ldots, T$ Time Periods
K independent variables

Note: $\exists$ no holes in the data:
For each i $\exists$ T obs. (time periods)
for each t $\exists$ N obs. (cross-sections)

In order to write this model as $Y = X\beta + \epsilon$, the data must be ordered;

$$Y = \left[\begin{array}{c} y_{11} \\ y_{12} \\ \vdots \\ y_{1T} \\ \hline y_{21} \\ y_{22} \\ \vdots \\ y_{2T} \\ \hline \vdots \\ \hline y_{N1} \\ y_{N2} \\ \vdots \\ y_{NT} \end{array}\right] \quad \epsilon = \left[\begin{array}{c} \epsilon_{11} \\ \epsilon_{12} \\ \vdots \\ \epsilon_{1T} \\ \hline \epsilon_{21} \\ \epsilon_{22} \\ \vdots \\ \epsilon_{2T} \\ \hline \vdots \\ \hline \epsilon_{N1} \\ \epsilon_{N2} \\ \vdots \\ \epsilon_{NT} \end{array}\right] \quad X = \left[\begin{array}{ccc} x_{111} & \cdots & x_{11K} \\ x_{121} & \cdots & x_{12K} \\ \vdots & & \\ x_{1T1} & \cdots & x_{1TK} \\ \hline x_{211} & \cdots & x_{21K} \\ x_{221} & \cdots & x_{22K} \\ \vdots & & \\ x_{2T1} & \cdots & x_{2TK} \\ \hline & \vdots & \\ \hline x_{N11} & \cdots & x_{N1K} \\ x_{N21} & \cdots & x_{N2K} \\ \vdots & & \\ x_{NT1} & \cdots & x_{NTK} \end{array}\right] \quad \beta = \left[\begin{array}{c} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_K \end{array}\right]$$

$NT \times 1$ $\qquad NT \times 1$ $\qquad NT \times K$ $\qquad K \times 1$

There are N blocks of T obs. each.

Motivation: To increase sample size.
By pooling, can turn small sample into a large sample.

e.g. — If I have 1 obs. on 10 firms, but can augment with obs. on each firm for 10 years; sample ↑es from 10 to 100.

— If I have obs. for 10 years on a particular state; if I can get the same obs. (time series) for the other 49 states, sample size ↑es from 10 to 500.

Note: Writing the model as $Y = X\beta + \epsilon$, presumes that ∃ only one set of β's for all cross-sections and for all time periods.

This is a very severe restriction!

* * * *

It is likely that the relationships are not identical accross-sections or over time!

We should test this restriction!

How can we test this restriction?

Consider: $y_{it} = x_{it}\beta + \overset{D_i}{\alpha_i} + \overset{D_t}{\delta_t} + \epsilon_{it}$

e.g. a single-input production function;
i^{th} firm, t^{th} time period.
$[i = 1, \dots, N; \quad t = 1, \dots, T]$

We expect β to be the same for all units;
but α_i may vary accross-sections,
and δ_t may vary accross time periods

$\alpha_i \Rightarrow$ marginal input
$\delta_t \Rightarrow$ technology

Since it is unrealistic to assume that all firms act exactly the same, allow α_i to measure the "constant effects associated with the i^{th} firm, but not attributable to any specific causal variables"

i.e. α_i is specific to i^{th} firm;
δ_t is specific to t^{th} period.

We may treat α_i and δ_t as:
(1) parameters to be measured, or as
(2) components of the error term.

(1) → Least Squares Dummy Variable (LSDV) Approach
(2) → Error Components Model.

(Fixed Effects)

(1) LSDV Approach

Here the intercept terms are treated as cross-section specific (α_i) or time specific (δ_t) Dummy Variables. (not and)

Suppose we have "unbalanced data"; $N \neq T$; the # of cross-sections is not close to the # of time periods.

e.g. Suppose T is large and N is small (long time series, for few cross-sections).

The Model:

$$y_{it} = \overset{D_i}{\alpha_i} + \beta_{i1} X_{it1} + \beta_{i2} X_{it2} + \cdots + \beta_{iK} X_{itK} + \epsilon_{it}$$

$$i = 1, \ldots, N; \quad t = 1, \ldots, T$$

OR: $$y_{it} = \overset{D_i}{\alpha_i} + \underset{(1\times K)}{X_{it}} \underset{(K\times 1)}{\beta_i} + \epsilon_{it} \qquad D_i = \begin{cases} 1 & \text{in } i^{th} \text{ cr-s} \\ 0 & \text{otherwise} \end{cases}$$

In any cross-section analysis, $\exists$ N obs. (given t) from which to estimate K β's.
In any particular time series, $\exists$ T obs. (given i) from which to estimate K β's.

With N small and T big, we would consider the pooled data as N different time series, where each time series has T obs. to estimate K β's in β_i, and α_i can measure the different intercepts accross sections.

Thus, order the data as follows.

$$\left.\begin{array}{l} y_{11} = \alpha_1 + X_{11}\beta_1 + \epsilon_{11} \\ \vdots \\ y_{1T} = \alpha_1 + X_{1T}\beta_1 + \epsilon_{1T} \end{array}\right\} \quad y_1 = \alpha_1 + X_1\beta_1 + \epsilon_1$$

$$\vdots \qquad\qquad\qquad\qquad \vdots$$

$$\left.\begin{array}{l} y_{N1} = \alpha_N + X_{N1}\beta_N + \epsilon_{N1} \\ \vdots \\ y_{NT} = \alpha_N + X_{NT}\beta_N + \epsilon_{NT} \end{array}\right\} \quad y_N = \alpha_N + X_N\beta_N + \epsilon_N$$

Written as "time series regressions on N different cross-sections

Like Seemingly Unrelated Regressions!

* We must be able to assume that $\beta_i = \beta_j$ for any $i, j = 1, \ldots, N$.

Otherwise (if $\beta_i \neq \beta_j$) we would have Unrelated Regressions, and it would not make sense to pool the data!

In Matrix Form,

$$\underset{NT\times 1}{\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_N \end{bmatrix}} = \underset{NT \times N(K+1)}{\begin{array}{c} \begin{array}{cccccccc} D_1 & D_2 & & D_N & & & & \end{array} \\ \begin{bmatrix} 1 & 0 & \cdots & 0 & X_1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 & 0 & X_2 & \cdots & 0 \\ \vdots & & & & & & & \\ 0 & 0 & \cdots & 1 & 0 & 0 & \cdots & X_N \end{bmatrix} \end{array}} \underset{N(K+1)\times 1}{\begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_N \\ \beta_1 \\ \beta_2 \\ \vdots \\ \beta_N \end{bmatrix}} + \underset{NT\times 1}{\begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_N \end{bmatrix}}$$

where y_i is $T\times 1$
1 is $T\times 1$ vector of ones
0 is $T\times 1$ vector of zeroes
X_i is $T\times K$
α_i is a scalar
β_i is $K\times 1$

Wish to test hypothesis that $\beta_i = \beta_j \quad \forall\, i,j$.

i.e.

$$H_0: \beta_1 = \beta_2 = \cdots = \beta_N$$

OR: $\Longrightarrow \begin{bmatrix} \beta_{11} = \beta_{21} = \cdots = \beta_{N1} \\ \beta_{12} = \beta_{22} = \cdots = \beta_{N2} \\ \vdots \\ \beta_{1K} = \beta_{2K} = \cdots = \beta_{NK} \end{bmatrix}$ The K different β's must be identical in all N cross-sections.

This amounts to NK linear restrictions.

Recall the general test statistic;

$$F = \frac{(SSE_R - SSE_u) / (\text{\# of restrictions})}{SSE_u / (\text{d.f. of unrestricted model})}$$

To obtain SSE_u, run the N different "time series regressions" separately;

$$y_i = \overset{D_i}{\alpha_i} + X_i \beta_i + \epsilon_i \qquad i = 1, \ldots, N$$

Then $SSE_u = SSE_1 + SSE_2 + \cdots + SSE_N$

and its d.f. $= \sum_{i=1}^{N} (T - (K+1)) = N(T-K-1)$.

To obtain SSE_R, simply pool the data and run the joint equation;

$$y_{it} = \overset{D_i}{\alpha_i} + X_{it}\beta + \epsilon_{it} \qquad i = 1, \ldots, N \; ; \; t = 1, \ldots, T$$

OR:

$$\underset{NT\times 1}{\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_N \end{bmatrix}} = \underset{NT\times(N+K)}{\begin{bmatrix} 1 & 0 & \cdots & 0 & X_1 \\ 0 & 1 & \cdots & 0 & X_2 \\ & \vdots & & & \\ 0 & 0 & \cdots & 1 & X_N \end{bmatrix}} \underset{(N+K)\times 1}{\begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_N \\ \beta \end{bmatrix}} + \underset{NT\times 1}{\begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_N \end{bmatrix}}$$

OR: $Y^* = X^* \beta^* + \epsilon^* \implies SSE_R$: d.f. $= NT-(K+N$

Thus, $$F = \frac{[SSE_R - (SSE_1 + SSE_2 + \cdots + SSE_N)] / (NK)}{(SSE_1 + SSE_2 + \cdots + SSE_N) / N(T-K-1)}$$

If F is large, reject H_0,
and pooling the data to form one model
is <u>not</u> appropriate.

If F is small, infer that the β_i do not
significantly differ from each other,
and it is OK to Pool the data.

<u>Comments</u>:

1. The α_i represent the variations (c-s fixed effects) (in the intercept) due to differences accross sections, but not to any causal variable.

2. The "balance" of the data should determine whether you estimate cross-sectional dummies (α_i) or time series dummies (δ_t) or both.

   e.g. if you have obs. on 5,000 families for 3 time periods, it would make sense to estimate dummies for time only (δ_t)!

   If you have 200 monthly obs. for 10 firms, estimate firm dummies only (α_i)!

3. The more dummies you estimate, the more degrees of freedom you lose.

e.g. (a) if you estimate (N) cross-sectional dummies, then

$$s_R^2 = \frac{SSE_R}{NT-(K+N)} = \frac{SSE_R}{N(T-1)-K}$$

(b) if you estimate (T) time dummies, then

$$s_R^2 = \frac{SSE_R}{NT-(K+T)} = \frac{SSE_R}{T(N-1)-K}$$

//If $N = 5000$, $T = 3$; a) $N(T-1) \Rightarrow 10000 - K = d.f.$
b) $T(N-1) \Rightarrow 14997 - K = d.f.$

So estimate Time dummies!

//If $N = 10$, $T = 200$; a) $N(T-1) \Rightarrow 1990 - K = d.f.$
b) $T(N-1) \Rightarrow 1800 - K = d.f.$

So estimate firm dummies!

If the data is "balanced," $(N \approx T)$ you may wish to consider another approach.

Error Components Model

Consider the effects specific to all cross-sections and specific to all time periods, as different components of the error term.

$$y_{it} = x_{it}\beta + \epsilon_{it} \qquad i=1,\dots,N; \quad t=1,\dots,T$$

where $\epsilon_{it} = u_i + v_t + w_{it}$

with $u_i \sim$ iid $N(0,\sigma_u^2)$ wrt cross-sections $i=1,\dots,N$

$v_t \sim$ iid $N(0,\sigma_v^2)$ wrt time periods $t=1,\dots,T$

$w_{it} \sim$ iid $N(0,\sigma_w^2)$ wrt all pooled data $(i=1,\dots,N;\ t=1,\dots,T)$

If ∃ no disturbance effects specific to all cross-sections or all time periods, $u_i = v_t = 0$, and $\epsilon_{it} = w_{it} \sim$ iid $N(0,\sigma_w^2)$.

In this case, the ideal conditions would be satisfied, and OLS estimates of $\hat{\beta}$ would be good.

However, as discussed earlier, ∃ likely to be heteroskedasticity across sections and autocorrelation over time. [$u_i, v_t \neq 0$]

The error components model is somewhat more realistic than the assumption that $\epsilon_{it} =$

Writing out the Model, $Y = X\beta + \epsilon$

$$\underset{NT\times 1}{Y} = \left[\begin{array}{c} y_{11} \\ y_{12} \\ \vdots \\ y_{1T} \\ \hline y_{21} \\ y_{22} \\ \vdots \\ y_{2T} \\ \hline \vdots \\ \hline y_{N1} \\ y_{N2} \\ \vdots \\ y_{NT} \end{array}\right] \quad \underset{NT\times K}{X} = \left[\begin{array}{c} X_{111}\ X_{112} \cdots X_{11K} \\ X_{121}\ X_{122} \cdots X_{12K} \\ \vdots \\ X_{1T1}\ X_{1T2} \cdots X_{1TK} \\ \hline X_{211}\ X_{212} \cdots X_{21K} \\ X_{221}\ X_{222} \cdots X_{22K} \\ \vdots \\ X_{2T1}\ X_{2T2} \cdots X_{2TK} \\ \hline \vdots \\ \hline X_{N11}\ X_{N12} \cdots X_{N1K} \\ X_{N21}\ X_{N22} \cdots X_{N2K} \\ \vdots \\ X_{NT1}\ X_{NT2} \cdots X_{NTK} \end{array}\right] \quad \underset{K\times 1}{\beta} = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_K \end{bmatrix} \quad \epsilon = \left[\begin{array}{c} \epsilon_{111} \\ \epsilon_{12} \\ \vdots \\ \epsilon_{1T} \\ \hline \epsilon_{21} \\ \epsilon_{22} \\ \vdots \\ \epsilon_{2T} \\ \hline \vdots \\ \hline \epsilon_{N1} \\ \epsilon_{N2} \\ \vdots \\ \epsilon_{NT} \end{array}\right]$$

$$E(\epsilon_{it}) = E(u_i) + E(v_t) + E(w_{it}) = 0$$

$$Cov(\epsilon_{it}) = Cov(u_i) + Cov(v_t) + Cov(w_{it})$$

since u_i, v_t, and w_{it} are independent.

Comments:

(1) This is like having random cross-section (u_i) and random time series (v_t) effects.

⇨ (2) Instead of estimating $N-1$ cross-section dummies (α_i) and $T-1$ time series dummies (δ_t),
estimate their variances, s_u^2 and s_v^2.

What does Cov(ε) look like?

$$\varepsilon = \begin{bmatrix} \varepsilon_{11} \\ \varepsilon_{12} \\ \vdots \\ \varepsilon_{1T} \\ \hline \varepsilon_{21} \\ \varepsilon_{22} \\ \vdots \\ \varepsilon_{2T} \\ \hline \vdots \\ \hline \varepsilon_{N1} \\ \varepsilon_{N2} \\ \vdots \\ \varepsilon_{NT} \end{bmatrix} \quad u = \underset{\text{cross-section}}{\begin{bmatrix} u_1 \\ u_1 \\ \vdots \\ u_1 \\ \hline u_2 \\ u_2 \\ \vdots \\ u_2 \\ \hline \vdots \\ \hline u_N \\ u_N \\ \vdots \\ u_N \end{bmatrix}} \quad v = \underset{\text{time period}}{\begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_T \\ \hline v_1 \\ v_2 \\ \vdots \\ v_T \\ \hline \vdots \\ \hline v_1 \\ v_2 \\ \vdots \\ v_T \end{bmatrix}} \quad \omega = \begin{bmatrix} \omega_{11} \\ \omega_{12} \\ \vdots \\ \omega_{1T} \\ \hline \omega_{21} \\ \omega_{22} \\ \vdots \\ \omega_{2T} \\ \hline \vdots \\ \hline \omega_{N1} \\ \omega_{N2} \\ \vdots \\ \omega_{NT} \end{bmatrix} \quad (\text{all } NT \times 1)$$

$$\mathrm{Cov}(\varepsilon) = \mathrm{Cov}(u) + \mathrm{Cov}(v) + \mathrm{Cov}(\omega)$$

since u, v, and ω are independent.

a) $\mathrm{Cov}(\omega) = E(\omega\omega') = \sigma_\omega^2 I_{NT}$

given assumption that $\omega_{it} \sim \text{iid } N(0, \sigma_\omega^2)$.

Since ω is indexed accross all sections and all time periods, each ω_{it} is independent from all others.

b) $\mathrm{Cov}(u) = E(uu')$

u represents randomness in the data that is specific to each of the N cross-sections regardless of the time period

Since u is indexed only accross sections, these disturbances are *not* independent accross different time periods.

$$\operatorname{Cov}(u) = E(uu') = E \left[\begin{array}{c} u_1 \\ \vdots \\ u_1 \\ \hline \vdots \\ \hline u_N \\ \vdots \\ u_N \end{array}\right] \left[u_1 \cdots u_1 \mid \cdots \mid u_N \cdots u_N\right]$$

$$= E \left[\begin{array}{ccc|ccc|c|ccc} u_1u_1 & \cdots & u_1u_1 & u_1u_2 & \cdots & u_1u_2 & & u_1u_N & \cdots & u_1u_N \\ \vdots & & & \vdots & & & \cdots & \vdots & & \\ u_1u_1 & \cdots & u_1u_1 & u_1u_2 & \cdots & u_1u_2 & & u_1u_N & \cdots & u_1u_N \\ \hline u_2u_1 & \cdots & u_2u_1 & u_2u_2 & \cdots & u_2u_2 & & u_2u_N & \cdots & u_2u_N \\ \vdots & & & \vdots & & & \cdots & \vdots & & \\ u_2u_1 & \cdots & u_2u_1 & u_2u_2 & \cdots & u_2u_2 & & u_2u_N & \cdots & u_2u_N \\ \hline & & & & \vdots & & & & & \\ \hline u_Nu_1 & \cdots & u_Nu_1 & u_Nu_2 & \cdots & u_Nu_2 & & u_Nu_N & \cdots & u_Nu_N \\ \vdots & & & \vdots & & & \cdots & \vdots & & \\ u_Nu_1 & \cdots & u_Nu_1 & u_Nu_2 & \cdots & u_Nu_2 & & u_Nu_N & \cdots & u_Nu_N \end{array}\right]$$

$$= \sigma_u^2 \begin{bmatrix} J_T & 0 & & 0 \\ 0 & J_T & & 0 \\ & & \ddots & \\ 0 & 0 & & J_T \end{bmatrix}$$

where J_T is a $T \times T$ matrix of *ones*.

Blocks of zeroes *off* the main diagonal indicate indep. between different cross-sections. Blocks of ones indicate correlatic among different time periods (in sa cr-se

$$= \sigma_u^2 (I_N \otimes J_T)$$

$$(NT \times NT)$$

$$\underline{(c)\ \mathrm{Cov}(v) = E(vv')}$$

$$= E \begin{bmatrix} v_1 \\ \vdots \\ v_T \\ \vdots \\ v_1 \\ \vdots \\ v_T \end{bmatrix} \left[v_1 \cdots v_T \mid \cdots \mid v_1 \cdots v_T \right]$$

$$= E \left[\begin{array}{c|c|c|c} v_1v_1 \cdots v_1v_T & v_1v_1 \cdots v_1v_T & & v_1v_1 \cdots v_1v_T \\ \vdots & \vdots & \cdots & \vdots \\ \underline{v_Tv_1 \cdots v_Tv_T} & \underline{v_Tv_1 \cdots v_Tv_T} & & \underline{v_Tv_1 \cdots v_Tv_T} \\ \vdots & & & \\ \hline v_1v_1 \cdots v_1v_T & v_1v_1 \cdots v_1v_T & & v_1v_1 \cdots v_1v_T \\ \vdots & \vdots & \cdots & \vdots \\ v_Tv_1 \cdots v_Tv_T & v_Tv_1 \cdots v_Tv_T & & v_Tv_1 \cdots v_Tv_T \end{array} \right]$$

$$= \sigma_v^2 \begin{bmatrix} I_T & I_T & \cdots & I_T \\ I_T & I_T & \cdots & I_T \\ \vdots & & & \\ I_T & I_T & \cdots & I_T \end{bmatrix} = \sigma_v^2 \; J_N \otimes I_T$$

$$(NT \times NT)$$

These Blocks, on and off the main diagonal, have <u>ones</u> on their own diagonals, indicating correlation among disturbances in the <u>same</u> time period (but in different cross-sections), and zeroes off their own diagonals, reflecting independence among disturbances in <u>different</u> time periods.

Since v is indexed only over time, these disturbances are <u>not</u> independent accross sections, if they are from the same time period.

Thus, $\mathrm{Cov}(\epsilon) = \sigma_w^2 I_{NT} + \sigma_u^2 (I_N \otimes J_T) + \sigma_v^2 (J_N \otimes I_T)$

$$= \sigma_w^2 I_{NT} + \sigma_u^2 A + \sigma_v^2 B$$

$$= \Omega .$$

$$= \begin{bmatrix} F & H & H & \cdots & H \\ H & F & H & \cdots & H \\ H & H & F & \cdots & H \\ & \vdots & & & \\ H & H & H & \cdots & F \end{bmatrix}$$

where $H = \begin{bmatrix} \sigma_v^2 & 0 & & 0 \\ 0 & \sigma_v^2 & & 0 \\ & & \ddots & \\ 0 & 0 & & \sigma_v^2 \end{bmatrix}$

and $F = \begin{bmatrix} (\sigma_u^2 + \sigma_v^2 + \sigma_w^2) & \sigma_u^2 & \cdots & \sigma_u^2 \\ \sigma_u^2 & (\sigma_u^2 + \sigma_v^2 + \sigma_w^2) & \cdots & \sigma_u^2 \\ \vdots & \vdots & & \vdots \\ \sigma_u^2 & \sigma_u^2 & \cdots & (\sigma_u^2 + \sigma_v^2 + \sigma_w^2) \end{bmatrix}$

Obviously, $\mathrm{Cov}(\epsilon) = \Omega \neq \sigma^2 I_{NT}$;

the ideal conditions are violated.

→ Use GLS!

$$Y = X\beta + \epsilon \qquad \text{with} \quad Cov(\epsilon) = \Omega$$

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

(see Wallace & Hussain, ECON, 1969)

* where $\Omega^{-1} = \frac{1}{\sigma_w^2}\left[I_{NT} - \gamma_1 A - \gamma_2 B + \gamma_3 J_{NT}\right]$

with $\gamma_1 = \frac{\sigma_u^2}{\sigma_w^2 + T\sigma_u^2}$, $\gamma_2 = \frac{\sigma_v^2}{\sigma_w^2 + N\sigma_v^2}$,

and $\gamma_3 = \gamma_1\gamma_2\left[\frac{2\sigma_w^2 + N\sigma_v^2 + T\sigma_u^2}{\sigma_w^2 + N\sigma_v^2 + T\sigma_u^2}\right]$

(verify yourself! $\Omega\Omega^{-1} = I_{NT}$)

If σ_u^2, σ_v^2, and σ_w^2 were known; γ_1, γ_2, and γ_3 would be known, and

$$\tilde{\beta}_{GLS} \sim N(\beta, (X'\Omega^{-1}X)^{-1}),$$

$$\sqrt{NT}(\tilde{\beta}_{GLS} - \beta) \rightarrow N\left(0, \lim_{NT\to\infty}\left(\frac{X'\Omega^{-1}X}{NT}\right)^{-1}\right)$$

Notes:

① N, T, and NT must each approach ∞. (more later [p.17 & 18])

② In this case, we wouldn't care about the asymptotic distribution, since we would know the small sample distribution!
— written for comparative purposes.

Consider the asymptotic covariance matrix.

$$\lim_{NT\to\infty}\left(\frac{X'\Omega^{-1}X}{NT}\right) = \lim_{NT\to\infty}\frac{1}{\sigma_w^2}\left[\frac{1}{NT}\left(X'I_{NT}X - \gamma_1 X'AX - \gamma_2 X'BX + \gamma_3 X'J_{NT}\right.\right.$$

(i) $\lim\limits_{NT\to\infty}\frac{X'X}{NT}$ exists

(ii) $\lim\limits_{NT\to\infty}\gamma_1\frac{X'AX}{NT} = \lim\limits_{NT\to\infty}(T\gamma_1)\frac{X'AX}{NT^2}$ is finite & nonsingul

$$\left[\text{and } \lim_{T\to\infty}(T\gamma_1) = \lim_{T\to\infty}(T)\frac{\sigma_u^2}{\sigma_w^2+T\sigma_u^2} = 1\right]$$

Observe that $x'Ax = x'\begin{bmatrix} J_T & 0 & 0 \\ 0 & J_T & 0 \\ & & \ddots \\ 0 & 0 & J_T \end{bmatrix}x$ is $K\times K$;

A has N blocks of T^2 ones; NT^2 nonzero elements. $X'AX$ has K^2 elements; each is a double sum over NT values; thus N^2T^2 values figure into each element of $X'AX$.

BUT, many elements in A are zero; (all the off-diagonal $T\times T$ blocks).

Thus, we actually eliminate many of these N^2T^2 terms; each element in $X'AX$ can be written as being indexed over T twice and N once.

$\therefore$ Actually, NT^2 values figure into each element of $X'AX$.

Hence, the appropriate index for $X'AX$, to get a finite nonsingular limit, is NT^2 rather than N^2T^2 or NT.

(iii) $\lim\limits_{NT\to\infty} \gamma_2 \frac{X'BX}{NT} = \lim\limits_{NT\to\infty} (N\gamma_2) \frac{X'BX}{N^2T}$ is finite & nons

$$\left[\lim_{N\to\infty}(N\gamma_2) = \lim_{N\to\infty}(N)\frac{\sigma_v^2}{\sigma_w^2 + N\sigma_v^2} = 1\right]$$

Observe that $X'BX = X' \begin{bmatrix} I_T & I_T & \dots & I_T \\ I_T & I_T & & I_T \\ \vdots & & & \\ I_T & I_T & \dots & I_T \end{bmatrix} X$ is $K\times K$.

B has N^2 blocks with T ones in each; N^2T nonzero elements.

Here the zeroes in B eliminate many terms in the double sums making up the K^2 elements of $X'BX$.

Thus actually, N^2T values really figure into each element of $X'BX$.

(iv) $\lim\limits_{NT\to\infty} \gamma_3 \frac{X'J_{NT}X}{NT} = \lim\limits_{NT\to\infty} (NT\gamma_3) \frac{X'J_{NT}X}{N^2T^2}$ is finite & nonsingular.

$$\lim_{N,T\to\infty}(NT\gamma_3) = \lim_{NT\to\infty} \underset{(1)}{\underset{\downarrow}{(T\gamma_1)}} \underset{(1)}{\underset{\downarrow}{(N\gamma_2)}} \underset{(1)}{\underset{\downarrow}{\left[\frac{2\sigma_w^2 + N\sigma_v^2 + T\sigma_u^2}{\sigma_w^2 + N\sigma_v^2 + T\sigma_u^2}\right]}} = 1$$

Here, J_{NT} has N^2T^2 ones; $X'J_{NT}X = X\begin{bmatrix} 111 \cdots 1 \\ 111 \cdots 1 \\ \vdots \\ 111 \cdots 1 \end{bmatrix} X$.

Since $\exists$ no zeroes in J_{NT}, all N^2T^2 values figure into each element of $X'J_{NT}X$, and N^2T^2 is the appropriate index.

ESTIMATION of the Error Components Model

1.) OLS (Bad)

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

$$\hat{\beta} \sim N\left[\beta, (X'X)^{-1}X'\Omega X(X'X)^{-1}\right]$$

$$\sqrt{NT}(\hat{\beta}-\beta) \rightarrow N\left[0, \lim_{NT\to\infty} NT(X'X)^{-1}X'\Omega X(X'X)^{-1}\right]$$

$$\text{or} \quad N\left[0, \lim_{NT\to\infty} \left(\frac{X'X}{NT}\right)^{-1} \frac{X'\Omega X}{NT} \left(\frac{X'X}{NT}\right)^{-1}\right]$$

Here, $\left(\frac{X'X}{NT}\right)^{-1}$ has a finite, nonsingular limit.

However, $\frac{X'\Omega X}{NT} = \sigma_w^2 \frac{X'X}{NT} + \sigma_u^2 \underset{\Uparrow}{\frac{X'AX}{\underline{\underline{NT}}}} + \sigma_v^2 \underset{\Uparrow}{\frac{X'BX}{\underline{\underline{NT}}}} \longrightarrow \infty$

These are not the magic numbers to index these last two terms by. We need $\underline{\underline{N^2T}}$ and $\underline{\underline{NT^2}}$ respectively for convergence.

Thus OLS has an infinite asymptotic covariance matrix.

BAD!

→ inconsistent!

2.) GLS, variances $(\sigma_u^2, \sigma_v^2, \sigma_w^2)$ known.

$$\Omega = \sigma_u^2 A + \sigma_v^2 B + \sigma_\omega^2 I_{NT}$$

$$\Omega^{-1} = \frac{1}{\sigma_\omega^2}\left(I_{NT} - \gamma_1 A - \gamma_2 B + \gamma_3 J_{NT} \right)$$

$$\tilde{\beta} = (X'\Omega^{-1}X)^{-1} X'\Omega^{-1} Y$$ nice. (See pp. 16-18.)

3.) GLS, variances unknown.

$\sigma_u^2, \sigma_v^2, \sigma_w^2$, and $\therefore$ γ_1, γ_2, & γ_3 unknown.

* It is a real pain to estimate them.

Use an iterative procedure:

(i) Estimate $Y = X\beta + \epsilon$ by OLS, to obtain residuals, $e_{it} = y_{it} - X_{it}\hat{\beta}_{OLS}$ — not ~~and~~ consistent ~~but~~ inefficient ~~see~~ p. 19 !!

(ii) Estimate σ_u^2, σ_v^2, and σ_w^2;

$$\tilde{\sigma}_u^2 = \frac{1}{T}\left[\sum_{i=1}^{N} \frac{e_{i\cdot}^2}{T(N-1)} - \tilde{\sigma}_\omega^2 \right]$$

$$\tilde{\sigma}_v^2 = \frac{1}{N}\left[\sum_{t=1}^{T} \frac{e_{\cdot t}^2}{N(T-1)} - \tilde{\sigma}_\omega^2 \right]$$

$$\tilde{\sigma}_\omega^2 = \frac{1}{(N-1)(T-1)} \sum_{i=1}^{N} \sum_{t=1}^{T} \left(e_{it} - \frac{1}{T} e_{i\cdot} - \frac{1}{N} e_{\cdot t} \right)^2$$

where $e_{i\cdot} = e_{i1} + e_{i2} + \cdots + e_{iT}$ (i^{th} cross-secti

$e_{\cdot t} = e_{1t} + e_{2t} + \cdots + e_{Nt}$ (t^{th} time period)

$\lim \tilde{\sigma}^2 - \sigma^2 = 0$ for each of these.

(iii) $\hat{\Omega}^{-1} = \frac{1}{\tilde{\sigma}_w^2}\left[I_{NT} - \hat{\gamma}_1 A - \hat{\gamma}_2 B + \hat{\gamma}_3 J_{NT} \right]$

where $\tilde{\sigma}_u^2, \tilde{\sigma}_v^2$, and $\tilde{\sigma}_w^2 \rightarrow \hat{\gamma}_1, \hat{\gamma}_2$, and $\hat{\gamma}_3$.

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

Provided some conditions are satisfied, $\hat{\beta}_{GLS}$ is asymptotically unbiased, consistent, and asymptotically efficient.
(like all GLS estimators in this context)

Testing; $\frac{\hat{\beta}_i}{(X'\hat{\Omega}^{-1}X)^{-1}_{ii}} \rightarrow t_{T-K}$ $\left[\hat{\Omega}^{-1} = f(\tilde{\sigma}_w^2)\right]$

It turns out that γ_1, γ_2, & γ_3 don't matter asymptotically,
but we need to know $\tilde{\sigma}_w^2$ for testing.

This is difficult to compute,
and may not be very good for small samples.

Is there a simpler method ?

4) Covariance Estimator, $\tilde{\beta}_{cov}$.

The reason we need to "estimate" σ_u^2, σ_v^2, and σ_w^2, is that they determine γ_1, γ_2, and γ_3 which appear in Ω^{-1}.

Again, it turns out that γ_1, γ_2, and γ_3 <u>don't matter asymptotically</u> (recall pp. 16-18).

If they don't matter asymptotically, why estimate them?

Simply let $\gamma_1 = \frac{1}{T}$, $\gamma_2 = \frac{1}{N}$, and $\gamma_3 = \frac{1}{NT}$.

PROCEDURE:

Want to find a [simple] transformation matrix, Q, such that $Q'Q = \Omega^{-1}$

Let $$Q = I_{NT} - \underset{(\gamma_1)}{\frac{1}{T}} A - \underset{(\gamma_2)}{\frac{1}{N}} B + \underset{(\gamma_3)}{\frac{1}{NT}} J_{NT} = \hat{\Omega}^{-1}$$

∃ no unknowns here!

A, B, J_{NT} composed of ones & zeroes.

(i) Q is: symmetric, idempotent, of dimension $NT \times NT$, of rank $(N-1)\times(T-1)$,

erify self!

(ii) $Qu = Qv = 0$

$$\Longrightarrow Q'Q = Q = \hat{\Omega}^{-1}.$$

[illegible] (2)

$$Qu = I_{NT}(u) - \overset{①}{\frac{1}{T}(I_N \otimes J_T)(u)} - \overset{②}{\frac{1}{N}(J_N \otimes I_T)(u)} + \overset{③}{\frac{1}{NT}J_{NT}(u)} = ($$

①
$$\underset{(NT\times NT)\,(NT\times 1)}{(I_N \otimes J_T)(u)} = \begin{bmatrix} Tu_1 \\ Tu_1 \\ \vdots \\ Tu_1 \\ \hline Tu_2 \\ Tu_2 \\ \vdots \\ Tu_2 \\ \hline \vdots \\ \hline Tu_N \\ Tu_N \\ \vdots \\ Tu_N \end{bmatrix} \qquad \text{Thus } \frac{1}{T}(I_N \otimes J_T)(u) = u$$

②
$$\underset{(NT\times NT)\,(NT\times 1)}{(J_N \otimes I_T)u} = \begin{bmatrix} I_T & I_T & \cdots & I_T \\ \vdots & & & \\ I_T & I_T & \cdots & I_T \end{bmatrix} \begin{bmatrix} u_1 \\ u_T \\ \hline u_1 \\ u_2 \\ \vdots \\ u_2 \\ \hline \vdots \\ \hline u_N \\ \vdots \\ u_N \end{bmatrix} = \underset{NT\times 1}{\begin{bmatrix} \sum_{i=1}^{N} u_i \\ \Sigma u_i \\ \vdots \\ \Sigma u_i \end{bmatrix}}$$

$$\text{Thus } \frac{1}{N}(J_N \otimes I_T)(u) = \underset{NT\times 1}{\begin{bmatrix} \bar{u} \\ \bar{u} \\ \vdots \\ \bar{u} \end{bmatrix}} = \bar{u}$$

③
$$J_{NT}(u) = \begin{bmatrix} 1 & \cdots & \\ \vdots & & \\ & \cdots & \end{bmatrix} \begin{bmatrix} u_1 \\ u_1 \\ u_2 \\ \hline u_2 \\ \hline u_N \\ \vdots \\ u_N \end{bmatrix} = \begin{bmatrix} Tu_1 + Tu_2 + \cdots + Tu_N \\ Tu_1 + Tu_2 + \cdots + Tu_N \\ \vdots \end{bmatrix} = T\begin{bmatrix} \sum_{i=1}^{N} u_i \\ \Sigma u_i \\ \vdots \\ \Sigma u_i \end{bmatrix}$$

$$\text{Thus, } \frac{1}{NT} J_N(u) = \begin{bmatrix} \bar{u} \\ \bar{u} \\ \vdots \\ \vdots \end{bmatrix} = \bar{u}$$

Combining: $Qu = (u) - (u) - (\bar{u}) + (\bar{u}) = 0$

The Transformed System:

$$QY = QX\beta + Q\epsilon$$

$$\Longrightarrow QY = QX\beta + Qw \qquad \boxed{\begin{matrix}\text{since } Qu = Qv = 0 \\ Q\epsilon = Qw.\end{matrix}}$$

Qw satisfies the ideal conditions;
Run OLS on transformed system;

$$\tilde{\beta}_{cov} = (X'Q'QX)^{-1} X'Q'QY$$

$$= (X'QX)^{-1} X'QY$$

Q symmetric & idempote

$$= \text{GLS using } \hat{\Omega}^{-1} = Q$$

★ This is called the covariance estimator (almost GLS

Moments:

$$E(\tilde{\beta}_{cov}) = E\left[(X'QX)^{-1} X'Q(X\beta + \epsilon)\right] = \beta \quad ✓$$

$$\begin{aligned} Cov(\tilde{\beta}_{cov}) &= E\left[(\tilde{\beta}-\beta)(\tilde{\beta}-\beta)'\right] \\ &= E\left[(X'QX)^{-1} X'Q(ww')QX(X'QX)^{-1}\right] \\ &= \sigma_w^2 (X'QX)^{-1} \quad ✓ \end{aligned}$$

$\epsilon = w + u + v$
but $Q\epsilon = Qw + Qu + Qv = Qw$ only!

symptotically efficient.

Observe that as N, T → ∞, this estimator has the same asymptotic covariance matrix as $\tilde{\beta}_{GLS}$.

$$\lim_{NT\to\infty} \left(\frac{X'QX}{NT}\right)^{-1} = \left\{ \lim_{NT\to\infty} \left(\frac{X'X}{NT} - \frac{X'AX}{NT^2} - \frac{X'BX}{N^2T} + \frac{X'J_{NT}X}{N^2T^2} \right) \right\}^{-1}$$

by defn. of Q (see p.16, 17) ✓

$$\Longrightarrow \sqrt{NT}\,(\tilde{\beta}_{cov} - \beta) \rightarrow N\left[0, \sigma_w^2 \lim_{NT\to\infty} \left(\frac{X'QX}{NT}\right)^{-1}\right].$$

Summarizing methods of estimating the Error Components Model.

1) OLS ~ Bad, if error components model assumptions are accurate.

2) GLS, variances known

3) GLS, variances consistently estimated

4) Covariance Estimator (almost GLS).

COMMENTS:

(i) The asymptotic efficiency of the last three methods are equivalent, as both N and $T \to \infty$.

e.g.	Balanced Data	Unbalanced	Unbalanced
	$N = 50$	$N = 5000$	$N = 10$
	$T = 50$	$T = 3$	$T = 120$
	yes	no	no
	[use error comp. model]	[use LSDV here.]	

(ii If you accept the assumptions of the error components model for your particular data and problem, use the Covariance Estimator $\tilde{\beta}$!

Does the data justify the assumptions about the disturbance term in the Error Components Model?

$$\epsilon_{it} = u_i + v_t + \omega_{it}$$

$$\begin{aligned} Cov(\epsilon) &= Cov(u) + Cov(v) + Cov(\omega) \\ &= \sigma_u^2 (I_N \otimes J_T) + \sigma_v^2 (J_N \otimes I_T) + \sigma_\omega^2 I_{NT} \\ &= \Omega \end{aligned}$$

$$= \begin{bmatrix} F & H & H & \cdots & H \\ H & F & H & \cdots & H \\ H & H & F & \cdots & H \\ & \vdots & & & \\ H & H & H & \cdots & F \end{bmatrix} \qquad (NT \times NT)$$

where $H = \sigma_v^2 I_T$ and $F = \begin{bmatrix} (\sigma_u^2 + \sigma_v^2 + \sigma_\omega^2) & \sigma_u^2 & \cdots & \sigma_u^2 \\ \sigma_u^2 & (\sigma_u^2 + \sigma_v^2 + \sigma_\omega^2) & \cdots & \sigma_u^2 \\ \vdots & \vdots & & \\ \sigma_u^2 & \sigma_u^2 & \cdots & (\sigma_u^2 + \sigma_v^2 + \sigma_\omega^2) \end{bmatrix}$

(i) $Cov(\epsilon_{it}, \epsilon_{is}) = \sigma_u^2 \quad \forall\ t \neq s$ (same i)

→ Disturbances in the same cross-section (i) are correlated accross time ($t \neq s$) in the F blocks, (but off the main diagonal).

$\boxed{\star 1} \longrightarrow$ This implies <u>constant</u> <u>autocorrelation</u> <u>in</u> any cross-section over time.

i.e. $Cov(\epsilon_{it}, \epsilon_{it-1}) = Cov(\epsilon_{it}, \epsilon_{it-2}) = \dots = \sigma_u^2$

These effects that are specific to cross-sections are constant, no matter how many time perio they are apart. (unrealistic)

The off-diagonal elements in the F blocks are all σ_u^2; these blocks are not just banded matrices, as stationarity would require, but all bands are even equal! Thus the (constant) autocorrelation that exists does not diminish over time.

(ii) $Cov(\epsilon_{it}, \epsilon_{jt}) = \sigma_v^2 \quad \forall \; i \neq j$ (same t)

→ Disturbances in the same time period (t) are correlated accross sections ($i \neq j$) on the diagonals of the H blocks.

$\boxed{\star 2}$ → This implies homoskedasticity accross sections (in the same time pd.).
For all cross-sections, the same variance (σ_v^2) applies to observations in the same time pd.

These assumptions (particularly ★1 & ★2) make the error components model less realistic than one might desire!
→ may have diminishing autocorrelation, and heteroskedasticity

A more realistic assumption (Kmenta):

(I) $\epsilon_{it} = \rho_i \epsilon_{it-1} + u_{it}$; $i = 1, \ldots, N$ $(i=j)$

(II) where $E(u_{it}) = 0$, $E(u_{it}^2) = \sigma_{ii}$, and

$$Cov(u_{it}, u_{js}) = \begin{cases} 0 & \text{if } t \neq s \\ \sigma_{ij} & \text{if } \underbrace{t = s}_{\text{time ser.}} \end{cases} \quad \underbrace{(\text{for } i \neq j)}_{\text{cross-section}}$$

Explanation:

(I) → First order autocorrelation accross time within each cross-section (ρ_i).

(II) → Contemporaneous correlation accross sec... that allows for heteroskedasticity. (σ_{ii} may vary with i along the main diagonal. σ_{ij} may vary with i and j along the "diagonals" of the off-diagonal blocks.)

This structure appears as follows.

First,

$$Cov(\epsilon_{it}, \epsilon_{is}) = \frac{\sigma_{ii}}{1-\rho_i^2} \begin{bmatrix} 1 & \rho & \rho^2 & \cdots & \rho^{T-1} \\ \rho & 1 & \rho & \cdots & \rho^{T-2} \\ \rho^2 & \rho & 1 & \cdots & \rho^{T-3} \\ & \vdots & & & \\ \rho^{T-1} & \rho^{T-2} & \rho^{T-3} & \cdots & 1 \end{bmatrix} = \underset{(T\times T)}{D_i}$$

$t, s = 1, \ldots, T$

Here, $i = j$: the T×T Blocks along the main diag.

Second,

$$\operatorname{Cov}(\epsilon_{it}, \epsilon_{js}) = \sigma_{ij} I_T = \underset{(T \times T)}{D_{ij}}$$

$i \neq j$; $t,s = 1,\ldots,T$

These represent the blocks off the main diag.

Third, together,

$$\operatorname{Cov}(\epsilon) = \Omega = \underset{(NT \times NT)}{\begin{bmatrix} D_1 & D_{12} & \cdots & D_{1N} \\ D_{21} & D_2 & \cdots & D_{2N} \\ \vdots & & & \\ D_{N1} & D_{N2} & \cdots & D_N \end{bmatrix}}$$

Implications:

① Within the same cross-section there may be first order autocorrelation (D_i , $i = 1,\ldots,N$).

② ∄ autocorrelation among disturbances accross sections; ∃ only contemporaneous correlation accross sections.

③ There may be heteroskedasticity accross sections; along the main diagonal, σ_{ii} ; $i = 1,\ldots,N$, off the main diagonal, σ_{ij} ; $i,j = 1,\ldots,N$.

Similar Structure to Error Comp. Model: more Realistic!

Estimation of this (more realistic) Model

$\text{Cov}(\epsilon) = \Omega$; use GLS!

If ρ_i, σ_{ii}, and σ_{ij} were known,
Ω would be known, and we could apply GLS directly.
Of course, these are rarely "known."

We can estimate them in a step by step Cochrane Orcutt procedure:

(1) Apply OLS to $Y = X\beta + \epsilon$, to obtain Least Squares residuals, $e_{it} = y_{it} - X_{it}\hat{\beta}_{OLS}$

(2) $\hat{\rho}_i = \dfrac{\sum_{t=2}^{T} e_{it}\, e_{it-1}}{\sum_{t=1}^{T} e_{it}^2}$ [Regress e_{it} on e_{it-1}].

(3) Form the transformed system:

$$y_{it}^* = \beta_1 x_{it1}^* + \beta_2 x_{it2}^* + \cdots + \beta_K x_{itK}^* + u_{it} \qquad i=1,\ldots,N;\ t=2,\ldots$$

where $y_{it}^* = y_{it} - \hat{\rho}_i y_{it-1}$; $x_{itj}^* = x_{itj} - \hat{\rho}_i x_{it-1j}$; and u_{it} is "nicer".
$\boxed{= \epsilon_{it} - \rho_i \epsilon_{it-1}}$

(4) Apply OLS to the transformed system $\rightarrow \tilde{\beta}_{GLS_1}$, to obtain these residuals;

$$\tilde{u}_{it} = y_{it}^* - x_{it}^* \tilde{\beta}_{GLS_1}$$

(5) From these GLS_1 residuals,

estimate the variance of u_{it} (σ_{ii}) as

$$\hat{\sigma}_{ii} = \frac{1}{T-K-1} \sum_{t=2}^{T} \tilde{u}_{it}^{2} \qquad i = 1, \ldots, N$$

ike SUR)

and the covariance between u_{it} and u_{jt} (σ_{ij}) as

$$\hat{\sigma}_{ij} = \frac{1}{T-K-1} \sum_{t=2}^{T} \tilde{u}_{it} \tilde{u}_{jt} \qquad i, j = 1, \ldots, N$$

(6) We now have estimates for D_i and D_{ij} $\Rightarrow \hat{\Omega}$.

Form GLS_2 with $\hat{\Omega}^{-1}$, to obtain $\boxed{\tilde{\beta}_{GLS_2}}$.

This will be asymptotically efficient.

[$\hat{\rho}_i$, $\hat{\sigma}_{ii}$, and $\hat{\sigma}_{ij}$ are consistent.]

Gory, but realistic!

Conclusion: Practically, we must take an eclectic approach. None of the specific methods [LSDV, Error Components, Kmenta extension] applies universally to all situations. Should use the model (re: the error structure) which approximates whats going on in your data the best.

* Should run simple OLS first, and look at the estimated intercepts and residuals.
If the intercepts behave systematically (rather than randomly) then the Least Squares Dummy Variable method would likely be appropriate to capture the implications. (with unbalanced data)
If the residuals don't exhibit any sort of strong systematic behavior (like heteroskedasticity or diminishin autocorrelation), then the restrictive assumptions of the error components model would be appropriate. However, if they do, then the error components model is inappropriate.

Of course Model choice is important, since the reliability of your estimates will depend on how closely the assumptions of the model truely apply to the data.

COMMENT
Given the restrictiveness of the Error Components Model, it is fairly amazing that so many people use it.