Beginning at the End

Tuesday, April 10, 2012
12:59 PM

This lecture is entitled "whatever you're doing with panel data, you're doing it wrong"

TSCS : time series cross section

Of the many problems that can exist in panel data, we will discuss:

- Spatial correlation of the errors *
- Contemporaneous correlation of the errors *
- Autocorrelation of the errors *
- Unit heterogeneity *

Spatial Correlation

Tuesday, April 10, 2012
1:34 PM

What it is: TSCS : N units (countries, people, etc.)

T time periods (years, months, etc.)

It might be the case that σ^2, the degree of noise in the DGP, is different for different units.

$$gdp_t = X_t \beta + u_t$$

Some countries more stable, less stable

Some countries have larger $var[u_t]$ than others.

$$E[u'u] \neq \sigma^2 I$$

the VCV of $(\hat{\beta}-\beta)(\hat{\beta}-\beta)^T \neq \sigma^2 (X'X)^{-1}$

t, F, etc. are incorrect. However, $\hat{\beta}$ is consistent

Consequences

$$\text{plim}_{T \to \infty} \frac{1}{\sqrt{T}} \sum_{t=1}^{T} a_t = 0.$$

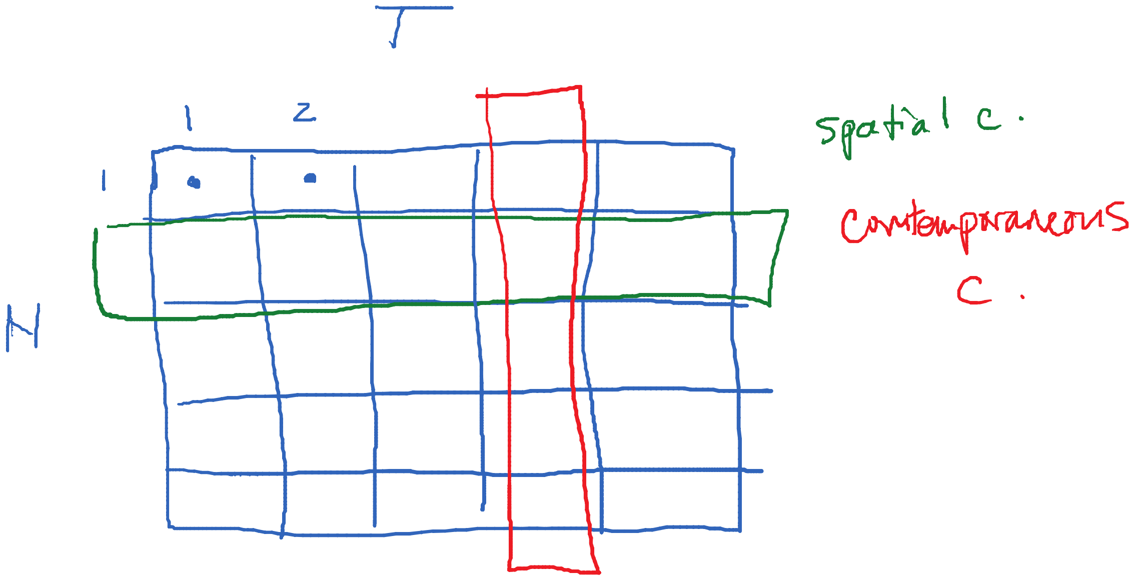
Difference between $\mu_k \neq \mu_j$ for 2 units k & j

$var[u_{kt}] \neq [u_{jt}]$ for 2 units k & j.

Contemporaneous Correlation

Tuesday, April 10, 2012
1:34 PM

What it is:



Consequences:

Same as spatial correlation

$$VCV(\hat{\beta}) \neq \sigma^2 (X'X)^{-1}.$$

$$uu^T \rightarrow \sigma^2 I$$

Panel-Corrected Standard Errors

Tuesday, April 10, 2012
1:34 PM

Uses the idea behind the Huber-White Heteroskedasticity-Consistent VCV matrix, but applying these ideas to the error structure implied by simultaneous spatial and contemporaneous correlation

$$VCV(\hat\beta) = [\hat\beta - \beta][\hat\beta - \beta]^T =$$

$$(X'X)^{-1} X' \underbrace{uu'}_{\Omega} X (X'X)^{-1}$$

$$uu' \to \sigma^2 I$$

$$(X'X)^{-1} X' \sigma^2 I X (X'X)^{-1}$$

$$\sigma^2 \cancel{(X'X)^{-1} X'X} (X'X)^{-1}$$

$$\sigma^2 (X'X)^{-1}.$$

std heteroskedasticity:

$$\Omega = \begin{bmatrix} \sigma_1^2 & 0 & 0 & 0 \\ 0 & \sigma_2^2 & 0 & 0 \\ 0 & & \ddots & 0 \\ 0 & 0 & & \sigma_n^2 \end{bmatrix} \longrightarrow \hat\Omega = \operatorname{diag}(\hat u \hat u^T)$$

$$\Omega = \begin{array}{c|c} & \begin{array}{cccc} t=1 & & & NT \end{array} \\ \begin{array}{c} t=1 \\ t=2 \\ t=3 \\ \vdots \\ t=T \end{array} & \begin{bmatrix} \Sigma_1 & 0 & 0 & \cdots & 0 \\ 0 & \Sigma_2 & 0 & \cdots & 0 \\ 0 & 0 & \Sigma_3 & & \vdots \\ \vdots & \vdots & & \ddots & \\ 0 & 0 & \cdots & & \Sigma_T \end{bmatrix} \end{array} NT$$

Σ_i = $N \times N$ matrix of observations for time t.

$$\Sigma_t = \begin{array}{c|c} & N \\ N & \begin{bmatrix} w_1^2 & w_1 & \cdots & w_{1N} \\ w_{21} & w_2^2 & \cdots & w_{2N} \\ \vdots & \vdots & \ddots & \\ w_{N1} & w_{N2} & \cdots & w_N^2 \end{bmatrix} \end{array}$$

matrix of observations for time t

$w_{ij} = cov(u_{it}, u_{jt})$

$w_i^2 = var(u_{it})$.

Properties of this VCV:

(1) all N units will have unique variances, w_i^2. These are the diagonal elements of Σ_i.

(2) the error for unit i at time t will be correlated with the error for unit j at time t. This is $\omega_i \omega_j = \omega_{ij}$

(3) The error for unit i at time t will <u>not</u> be correlated with the error for other units at other times — those are the blocks of zeros.

Estimation:

$$\omega_i^2 \rightarrow \frac{1}{T} \sum_{t=1}^{T} \hat{u}_{it}^2 = \hat{\omega}_i^2 = \widehat{var(u_i)}$$

$$w_{ij} = \omega_i \omega_j = \frac{1}{T} \sum_{t=1}^{T} \hat{u}_{it} \hat{u}_{jt} = \widehat{cov(u_i, u_j)}$$

$$\underset{T \times N}{E} \qquad \begin{array}{c|c} & 1 \qquad\qquad\qquad N \\ \begin{array}{c} 1 \\ \\ \\ T \end{array} & \begin{bmatrix} \hat{u}_{11} & \hat{u}_{12} & \cdots & \hat{u}_{1N} \\ \hat{u}_{12} & \hat{u}_{22} & \cdots & u_{2N} \\ \vdots & \vdots & \ddots & \hat{u}_{TN} \\ \hat{u}_{T1} & \hat{u}_{T2} & \cdots & \end{bmatrix} \end{array}$$

$$\underset{N \times N}{\hat{\Omega}} \quad E'E \cdot \frac{1}{T}$$

Note: CONSISTENT, not unbiased!

variances / covariances are calculated as averages of a unit over time. More N won't help w/ consistency.

We need $T \to \infty$ (technically) for consistency to apply.

Autocorrelation

Tuesday, April 10, 2012
1:34 PM

What it is: u_t is correlated with u_{t-1}, u_{t-2} ...

one common form: Autoregressive error with one lag, AR(1)

$$y_t = \gamma y_{t-1} + x\beta + u_t$$

if y_{t-1} is omitted,

$$y_t = x\beta + u_t \quad , \quad u_t = \rho u_{t-1} + \varepsilon$$

$$\rho \in [-1, 1]$$

Consequences: inefficient estimate of VCV. $\hat{\beta}$ is still consistent.

There are several ways to fix...

AC Corrections

Tuesday, April 10, 2012
1:39 PM

LDV model:

$$\hat{y}_t = \hat{\gamma}\, y_{t-1} + x\hat{\beta} + u \qquad (*)$$

degree of autocorrelation

PCSEs with AR(1) Correction:

try to build an $\hat{\Omega}$ in the VCV estimation that accounts for the degree of autocorrelation

Unit Heterogeneity

Tuesday, April 10, 2012
1:34 PM

What it is: Again, NOT fixed by PCSEs

$$y_t = X_t \beta + D_t \alpha + u_t$$

α_i: the effect of being in unit i

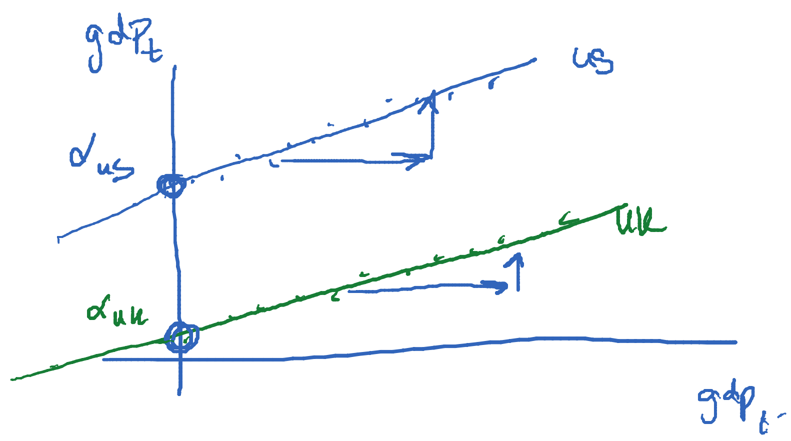
D_t: a matrix of N many dummy variables corresponding to the units

→

D_1	D_2	...	D_N
1	0		0
0	1		0
0	0		0
0	0		0
0	0		0

Consequences:

$\hat{\beta}$ can be biased due to OVB.



Fixed Effects

Tuesday, April 10, 2012
1:35 PM

Fancy name for "put in a lot of unit dummy variables"

$$y_t = X_t \beta + D\alpha + u$$

"fixed effects" = dummy variables

1. omit constant, put in N dummies
2. include constant, put in N-1 dummies

Advantages and Disadvantages:

+ this model is BLUE for a DGP w/ unit heterogeneity.

− this model is theoretically bereft.

+ easy to implement

- if elements of X are slow-moving over time (within units) dummy variables will be collinear w/ X and efficiency is harmed.

- if X is fixed, must drop X

- as the $N \to \infty$, $\hat{\beta}$ is not consistent. $D \to \infty$. This is the so-called "incidental parameters" problem.

Random Effects

Tuesday, April 10, 2012
1:35 PM

Add assumptions about the distribution of unit effects...

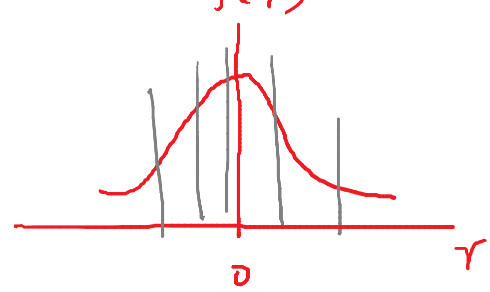
$$\mathrm{Var}(u_{it}) = \underbrace{\sigma^2_\gamma}_{\text{variance ascribable to unit heterogeneity}} + \underbrace{\sigma^2_\varepsilon}_{\text{normal error variance}}$$

$\mathrm{Cov}(u_{it}, u_{is}) = \sigma^2_\gamma$: within-unit error correlation is homogeneous

$\mathrm{Cov}(u_{it}, u_{js}) = 0$

the unit effects $\gamma \sim \mathrm{iid}(\mu = 0, \sigma^2 = \sigma^2_\gamma)$

each observation $y_{iT} = x_t\beta + \gamma_i + \varepsilon_{it}$



...if the assumptions are true, estimates are efficient!

$$\underset{NT \times NT}{\Omega} : \begin{bmatrix} \Sigma_i & 0 & 0 & \cdots & 0 \\ 0 & \Sigma_i & 0 & \cdots & 0 \\ 0 & 0 & \Sigma_i & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \\ 0 & 0 & 0 & & \Sigma_i \end{bmatrix}$$

$$\Sigma_i : \begin{array}{c|cccc} & t=1 & t=2 & \cdots & t=T \\ t=1 & \sigma^2_\varepsilon + \sigma^2_\gamma & \sigma^2_\gamma & \cdots & \sigma^2_\gamma \end{array}$$

$$\Sigma_i \quad \begin{array}{c} \\ T \times T \end{array} : \quad \begin{array}{c} t=1 \\ t=2 \\ \vdots \\ t=T \end{array} \begin{bmatrix} \sigma^2_\varepsilon + \sigma^2_\gamma & \sigma_\gamma^2 & \cdots & \sigma_\gamma^2 \\ \sigma_\gamma^2 & \sigma^2_\varepsilon + \sigma^2_\gamma & \cdots & \sigma_\gamma^2 \\ \vdots & & \ddots & \sigma^2_\varepsilon + \sigma^2_\gamma \\ \sigma_\gamma^2 & \sigma_\gamma^2 & \cdots & \end{bmatrix}$$

Now can't just correct SEs: we have OVB.

$E[u|X]=0$. We can do that by "weighting" the errors.

$$u'u = (y - X\hat{\beta})'(y - X\hat{\beta}) \rightarrow E[u/X] \neq 0.$$

$$\min \quad \hat{u}'\hat{\Omega}^{-1}\hat{u} = (y - X\hat{\beta})' \Omega^{-1} (y - X\hat{\beta})$$

$$\hat{\sigma}^2_\varepsilon = Var(\hat{\varepsilon}_i) = \frac{1}{NT-N-k}\sum_{i=1}^{NT} \hat{\varepsilon}_i^2 \qquad \hat{\sigma}_\gamma^2 = Var(\alpha) = \frac{1}{N}\sum_{i=1}^{N} \alpha_i$$

$$y_i = X\hat{\beta} + D\hat{\alpha} + \hat{\varepsilon}_i \qquad N \to \infty \text{ for a good estimates.}$$

Advantages and disadvantages

this is FGLS.

+ Can use slow-moving or fixed X.
+ if assumption u are true, then RE is more efficient than FE.

— if the assumptions are NOT true, $\hat{\beta}$ is biased.