Time-series–Cross Section Data: Modeling Cross-Sectional Issues

Nathaniel Beck (with a little help from his friends, particularly Jonathan Katz)

Department of Politics, NYU, New York, NY 10012, nathaniel.beck@nyu.edu

Prepared for Firenze Workshop, May 11, 2009

Overview

- How to analyze time-series–cross-sectional data
- Application to comparative politics (mostly political economy)
- Broadly defined, so includes most IR
- Excludes political behavior
- Today cross-sectional issues
- Next two sessions time-series issues
- PLEASE ASK QUESTIONS
- I may not answer them!
- Lots to cover, maybe too much

Outline

- Intro
- Panel vs TSCS data
- Notation
- Estimation issues lightly
- Specification
- Unit heterogeneity
  - Exploratory Analysis
  - Fixed Effects
  - Cross-validation
  - Random coefficients
- Spatial ideas

What is longitudinal data?

Observed over time as well as over space.

- Pure cross-section data has many limitations (Kramer, 1983). Problem is that only have one historical context.
- (Single) time series allows for multiple historical context, but for only one spatial location.
- Longitudinal data - repeated observations on units observed over time
- Subset of *hierarchical data* — observations that are correlated because there is some tie to same unit.
- E.g. in educational studies, where we observe student i in school u. Presumably there is some tie between the observations in the same school.
- In such data, observe $y_{j,u}$ where u indicates a unit and j indicates the j'th observation drawn from that unit. Thus no relationship between $y_{j,u}$ and $y_{j,u'}$ even though they have the same first subscript. In true longitudinal data, t represents comparable time.

Types of longitudinal data

- "Panel study" (NES, PSID, Congressional election outcomes by CD and year)
- Often use panel data as a single "enriched" cross-section, with info on prior behavior
- "Time-Series–Cross-Section" (political economy data on 15 OECD nations observed annually)
- Event history data
- Dyad year design in IR
- Data combining different surveys taken at different times (eg Markus article on Kramer)
- Rolling Cross-Section (Canadian Election Study)
- "Pseudo Panel" (group respondents by cohort) based on "Repeated Cross Section Data" (eg Family Expenditure Surveys)
- Binary dependent variable - estimation of transition matrix - Markov process

Panels vs TSCS data

- Logically TSCS data looks like panel data,
- but panels have large number of cross-sections (big N)
- with each unit observed only a few times (small T);
- TSCS data has reasonable sized T and not very large N
- For panel data, asymptotics are in N, T is fixed.
- For TSCS data, asymptotics in T, N is fixed.
- This distinction is critical
- Many of the panel methods are designed to deal with what is known as the "incidental parameters" problem, that is, as the number of parameters goes to ∞, one loses consistency
- This is a problem only for panel, not TSCS data.
- Furthermore, with small T there is no hope of saying anything about the time series structure of the data
- with "biggish" T there is.
- Care about the units in TSCS data; they are states or countries
- Not usually care about the units in panel models;

Issues that always arise in longitudinal data

- How model non-independent observations?
  - Repeated observations on same unit are seldom independent
  - Assumption of independence should be tested, not assumed
  - Next two sessions
- How model homogeneity?
  - Complete heterogeneity (area studies)
  - Complete homogeneity (econometrics)
  - Both positions silly, what is good compromise? Fixed effects?

Notation

The generic model

$$y_{i,t} = \mathbf{x}_{i,t}\boldsymbol{\beta} + \epsilon_{i,t}; \qquad i = 1, \ldots, N \qquad (1)$$
$$t = 1, \ldots, T$$

More details

- $\mathbf{x}_{i,t}$ is a K vector of exogenous variables
- observations are indexed by both unit (i) and time (t).
- $\boldsymbol{\Omega}$ is $NT \times NT$ covariance matrix of the errors
- typical element $E(\epsilon_{i,t}\epsilon_{j,s})$.
- Note: we are assuming a "rectangular" structure of the data;
- Not critical, but makes notation simpler
- ASSUME THAT y IS CONTINUOUS, NOT DISCRETE.
- While there is no real bounds on N, in typical applications it will be between 10 and 100.
- Assume T is large enough so that time averages make sense (say at least 10). In applications, T's of 20-50 are common

Estimation - spherical errors

- Assuming that the errors in Equation 1 are spherical
- (that is, satisfy the Gauss-Markov assumptions)
- OLS is optimal if the model is appropriately specified.
- The errors are spherical if all errors are independent and identically distributed so that

$$E(\epsilon_{i,t}\epsilon_{j,s}) = \begin{cases} \sigma^2 & \text{if } i = j \text{ and } s = t \\ 0 & \text{otherwise.} \end{cases} \tag{2}$$

- But even if we like spherical errors, Equation 1 ASSUMES:
  - all differences between units are accounted for by differences in the independent variables, that is, no "unmodeled heterogeneity"
  - no effects of other units on each other - no spatial effects
  - homogeneity (all units obey same equation) or "pooling"
  - no dynamics (temporal dependence)

Non-spherical errors

- Unlikely that cross-national panel errors will meet the assumption of sphericality.
- Usual OLS formula for standard errors will (always? often? sometimes?) provide misleading indications of the sampling variability
- The correct formula is given by the square roots of the diagonal terms of

$$\text{Cov}(\hat{\beta}) = (\mathbf{X}'\mathbf{X})^{-1}\{\mathbf{X}'\boldsymbol{\Omega}\mathbf{X}\}(\mathbf{X}'\mathbf{X})^{-1}. \quad (3)$$

- OLS estimates this by

$$\widehat{\text{Cov}(\hat{\beta})} = (\mathbf{X}'\mathbf{X})^{-1}\left\{\left(\frac{\sum_i \sum_t e_{i,t}^2}{NT-k}\right)\mathbf{X}'\mathbf{X}\right\}(\mathbf{X}'\mathbf{X})^{-1} \quad (4)$$

- which then simplifies to the usual OLS estimate of the variance-covariance matrix of the estimates

Correct VCV matrix - PCSEs

- OLS standard errors will be incorrect
- That is underestimate sampling variance of estimator
- If errors not spherical
- AND they structure of $\boldsymbol{\Omega}$ is related to the $X'X$ matrix
- PCSEs provide correct estimates of sampling variance of estimators
- A fixup, should model the issue, not fix it up
- This is what we do shortly
- Details below not critical

PCSE - Details

- Observations are stacked by unit (not time)
- First observation is unit 1, time 1, then unit 2, time 1, then unit 3, time 1, etc.
- VCV matrix for panel heteroskedastic and contemporaneously correlated (but temporally independent errors):

PCSEs

$$\boldsymbol{\Omega} = \begin{pmatrix} \boldsymbol{\Sigma} & \mathbf{0} & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Sigma} & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \boldsymbol{\Sigma} & \dots & \mathbf{0} \\ & & & . & \\ & & & . & \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \dots & \boldsymbol{\Sigma} \end{pmatrix} = \boldsymbol{\Sigma} \otimes \mathbf{I_N} \tag{5}$$

where

$$\boldsymbol{\Sigma} = \begin{pmatrix} \sigma_1^2 & \sigma_{1,2} & \sigma_{1,3} & \dots & \sigma_{1,N} \\ \sigma_{1,2} & \sigma_2^2 & \sigma_{2,3} & \dots & \sigma_{2,N} \\ & & & . & \\ & & & . & \\ \sigma_{1,N} & \sigma_{2,N} & \sigma_{3,N} & \dots & \sigma_N^2 \end{pmatrix} \tag{6}$$

PCSEs - estimating

- $\mathbf{\Omega}$ is an $NT \times NT$ matrix block diagonal matrix with
- $N \times N$ matrix of contemporaneous covariances, $\mathbf{\Sigma}$ (having typical element $E(\epsilon_{i,t}\epsilon_{j,t})$), along the diagonal
- To estimate Equation 4 need an estimate of $\mathbf{\Sigma}$
- OLS estimates of Equation 1 are consistent
- can use the OLS residuals from that estimation to estimate $\mathbf{\Sigma}$.
- Let $e_{i,t}$ be the OLS residual
- Estimate a typical element of $\mathbf{\Sigma}$ by

$$\hat{\mathbf{\Sigma}}_{i,j} = \frac{\sum_{t=1}^{T} e_{i,t} e_{j,t}}{T}. \tag{7}$$

- so estimate $\mathbf{\Sigma}$ by

$$\hat{\mathbf{\Sigma}} = \frac{(\mathbf{E}'\mathbf{E})}{T} \tag{8}$$

- and then estimate $\mathbf{\Omega}$ in the obvious way

PCSEs

- Compute "Panel Correct Standard Errors" (PCSEs) by taking the square root of the diagonal elements of

$$(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\left(\frac{\mathbf{E}'\mathbf{E}}{T} \otimes \mathbf{I}_T\right)\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}. \tag{9}$$

- Note: Different from Huber-White HCSEs
- HCSEs are only for (non-panel) heteroskedasticity

Generalized Least Squares

DO NOT USE FOR TSCS To Heterogeneity

- **Can estimate TSCS models with GLS**
- **If Ω is known (up to a scale factor), GLS is fully efficient and yields consistent estimates of the standard errors**
- **The GLS estimates of β are given by**

$$(\mathbf{X}'\mathbf{\Omega}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{\Omega}^{-1}\mathbf{Y} \tag{10}$$

- **with estimated covariance matrix**

$$(\mathbf{X}'\mathbf{\Omega}^{-1}\mathbf{X})^{-1}. \tag{11}$$

- **Usually just find a transform to do this (Cochrane-Orcutt for serial correlation)**

FEASIBLE GLS

- The problem is that $\mathbf{\Omega}$ is never known in practice (even up to a scale factor)
- An estimate of $\mathbf{\Omega}$, $\widehat{\mathbf{\Omega}}$, is used in Equations 10 and 11
- This procedure, FGLS, provides consistent estimates of β if
- $\widehat{\mathbf{\Omega}}$ is estimated by residuals computed from consistent estimates of β
- In finite samples FGLS underestimates sampling variability (for normal errors).
- This problem is not severe if there are only a small number of parameters in the variance-covariance matrix to be estimated (as in Cochrane-Orcutt)
- but is severe if there are a lot of parameters relative to the amount of data.

Panel Heteroskedasticity

- Panel het:

$$E(\epsilon_{i,t}\epsilon_{j,s}) = \begin{cases} \sigma_i^2 & \text{if } i = j \text{ and } s = t \\ 0 & \text{otherwise} \end{cases} \tag{12}$$

- GLS: estimate σ_i^2 from the residuals in the obvious way
- use in a weighted least squares procedure.
- The problem with this procedure is that it is basically weighting units by how well they fit the underlying regression
- Note how different this is from theoretically motivated weighted least squares.

Contemporaneously correlated errors

Errors are contemporaneously correlated if

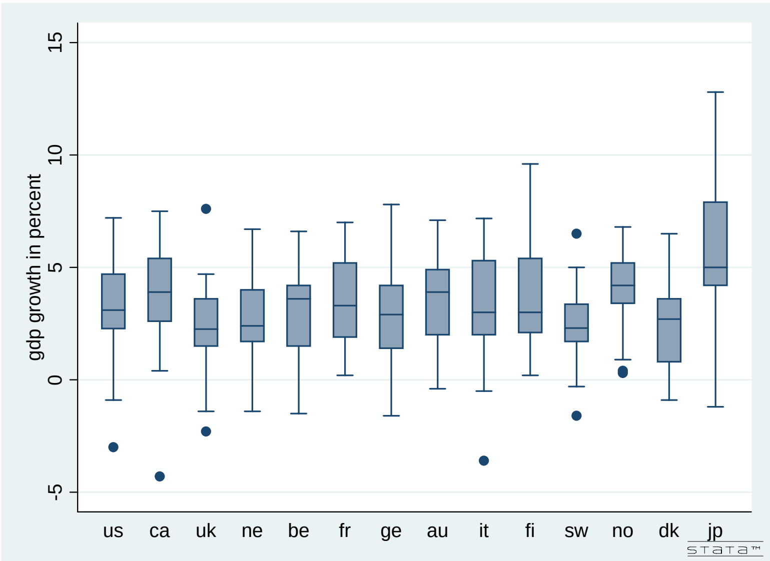
$$E(\epsilon_{i,t}\epsilon_{j,s}) = \begin{cases} \sigma_{i,j}^2 & \text{if } s = t \\ 0 & \text{otherwise.} \end{cases} \tag{13}$$

- (Kmenta)-Parks does FGLS
- Note the huge number of $\sigma_{i,j}^2$ to be estimated in GLS
- Horrible standard errors, off by 50% or more
- unless $T >> N$
- DO NOT USE KMENTA-PARKS (no one does these days)

Assessing Heterogeneity and simple fixes

- Good idea to look at plots, at least of dv by unit (country)
- Nice is box plot by unit
- For Garrett data, with depvar being unem, Stata command is
- `graph box unem, over(country)`
- Can also do by time

Example from Garrett data



Fixed Effects

- Equation 1 assumes that all countries are fit by same model.
- Easy fix is to simply adjoin to the equation country specific intercepts α_i.
- Just dummy variables added to the OLS
- Cause no estimation problems.
- Each effect is estimated with T observations, so no problem as long as $T \to \infty$, even with large (very large) N.

Testing for fixed effects

- Can test for whether need fixed effects by standard F-test
- Compare the SSEs in usual way between Equation 1
- and the specification with all the country dummies
- Some leave in only significant country dummies
- but that is probably less than best practice
- All the fixed effects do is shift each country's regression line up and down
- but leaves all regression lines parallel.
- If fixed effects are needed in the model and you exclude,
- you will have specification error (omitted variable bias)
- very serious if a) the unit effects are non-trivial
- and b) the unit variables are correlated with the x's in the model.

What it means to include FE's

- It is not benign to include FE's
- The standard results on partitioned regression show that
- the inclusion of FE's is equivalent to regressing the unit centered y's on the unit centered x's.
- Thus all the "between unit" variation in the variables is taken up by the FE's.
- Particular problem for TSCS data
- where the IV's may change very slowly
- If any does not change over time in any given unit
- it is completely colinear with the dummy variable for that country
- Return to more interesting heterogeneity later

A fix for time-invariant and near-time invariant iv's with FE

- Due to Plümper and Troeger (Political Analysis)
- For time-invariant iv's, just regress FEs on the the variables, then include
- Both time-invariant variables and the residuals (so same as FEs) in the model
- For near time-invariant, do the same
- Induces bias, but much lower variance, so often better MSE properties

Cross-Validation

- Which countries do not fit the pooled specification.
- Use cross-validation (c-v)
- In simplest form, c-v estimates a model leaving out one obs at a time, and then compares the prediction for that obs with the actual obs.
- Cross-sectional analogue of out-of-sample time series forecasting.
- For large data sets, cross-validated errors converge on the residuals
- Dropping one obs hardly changes the estimated parms
- For small data sets simple c-v is quite useful.

C-V for TSCS

- Can also do "leave-out-k(%)" as well as "leave-out-one" c-v.
- works nicely for TSCS data
- Just leave one country out at a time, predict it, and then examine the "forecast" errors.
- This is done in the table for a model of Garrett's
- Typical mean absolute forecast errors range from 1.2 to 2 (the unit is percent growth in GDP), except for Japan, which has a forecast error of 3.2% of GDP
- Thus clearly Japan fits the basic specification much less well than any other OECD nation.
- Done with `xvgarrett.do`

MAE of c-v errors by country for economic growth in 14 OECD nations, 1966–1990

Country	Mean absolute error
US	1.9
Canada	1.7
UK	1.7
Netherlands	1.6
Belgium	1.6
France	1.2
Germany	1.4
Austria	1.3
Italy	1.7
Finland	2.0
Sweden	1.2
Norway	1.5
Denmark	1.7
Japan	3.2

Heterogeneity of more than intercepts

- The classic approach to test for pooling
- Equation 1 as the null
- alternative being complete heterogeneity

$$Ho : y_{i,t} = \mathbf{x}_{i,t}\boldsymbol{\beta} + \epsilon_{i,t} \tag{14}$$
$$H1 : y_{i,t} = \mathbf{x}_{i,t}\boldsymbol{\beta}_i + \epsilon_{i,t} \tag{15}$$

- $H0 : \boldsymbol{\beta}_i = \boldsymbol{\beta}$.
- standard F test
- $k = 5$ and $N = 20$, 100 df in the numerator, a lot.

Random coefficient models

- Can estimate a model for each unit separately,

$$y_{i,t} = \mathbf{x}_{i,t}\boldsymbol{\beta}_i + \varepsilon_{i,t} \tag{16}$$

- If T is large enough it is not ridiculous to estimate N separate time-series
- but need at a BARE MINIMUM $T > 50$
- unusual for TSCS data

RCM (Mixed)

- Nice compromise is the "random coefficients model" (RCM)
- allows for unit heterogeneity
- but assumes that the various unit level coefficients are draws from a common (normal) distribution
- Thus the RCM adjoins to Equation 16

$$\boldsymbol{\beta}_i \sim N(\boldsymbol{\beta}, \boldsymbol{\Gamma}) \tag{17}$$

- where $\boldsymbol{\Gamma}$ is a matrix of variance and covariance terms to be estimated
- $\boldsymbol{\Gamma}$ indicates the degree the heterogeneity of the unit parameters
- $\boldsymbol{\Gamma} = \mathbf{0}$ indicates perfect homogeneity
- An important (and restrictive) assumption is that the stochastic process which generates the $\boldsymbol{\beta}_i$
  - is independent of the error process
  - and is uncorrelated with the vector of independent variables
  - These assumption less restrictive than the assumption of complete homogeneity.

Estimating RCMs

- Often estimated by Bayesian methods
- Feasible to estimate it via standard maximum likelihood
- The Pinheiro and Bates routines are implemented in R (`lmer`)
- Can also use Stata `xtmixed`
- Note: get "predictions of the $\boldsymbol{\beta}_i$
- Looking at the estimates of the variance terms may tell us to fix some parameters rather than leave them random
- Can either assume a general $\boldsymbol{\Gamma}$ (so randomness covaries)
- Probably "too flexible'," so can assume $\boldsymbol{\Gamma}$ is diagonal
- (no covariances of randomness)

Modeling the heterogeneity

- One especially nice feature of the RCM (inherited from the multilevel model)
- Can model the variation of the unit coefficients as a function of unit level variables
- T he marginal effect of some $x_{i,t}$ on $y_{i,t}$ can be made dependent on some unit level z_i
- Can make Equation 17 more general by

$$\boldsymbol{\beta}_i = \boldsymbol{\alpha} + \mathbf{z}_i \boldsymbol{\kappa} + \boldsymbol{\mu}_i \tag{18}$$

- where the μ are drawn from k-variate normal distribution
- This just produces an RCM with interaction terms
- Interaction terms tell us effect of left govt is dependent on labor organization

Spatial ideas - joint with Kristian Gleditsch

- There may be some relationship between units
- with a bigger relationship between "nearby" units
- Could be in errors ("spatially lagged errors")

$$y_{i,t} = \mathbf{x}_{i,t}\boldsymbol{\beta} + \epsilon_{i,t} + \sum_{j \neq i} w_j \epsilon_{j,t} \tag{19}$$

- where the w_j are the measures of "distance"
- Would solve FGLS problem, since now only one extra parm

- Or could enter as spatial lag (another ind var)

$$y_{i,t} = \mathbf{x}_{i,t}\boldsymbol{\beta} + \epsilon_{i,t} + \sum_{j \neq i} w_j y_{j,t-1} \tag{20}$$

- Note the temporal lag in the spatial lag here
- This makes estimation easy but assumption is critical
- Also must assume that errors are temporally independent (TUESDAY!)
- This is Garrett's OECD demand variable

“Distance”

- In geographical models, distance is easy idea (Euclidean distance or contiguity)
- But in our models can use any numbers that are non-negative to give idea of “political or economic” distance
- Could be “effective distance” (how close are India and Pakistan?)
- In political economy, two nations are “close” if they trade a lot
- For democracy, could be flow of communications