POL 606 Time Series Analysis
Prof. Matthew Lebo
Weeks 12-14 2008
Pooled Cross-Sectional Time Series Analysis

Introduction[1]

What is panel data / pooled time series cross-sectional analysis?

Data that follows a given sample of units (e.g. individuals, countries, dyads), $i=1,2,\ldots N$ over time, $t=1,2,\ldots,T$, so that we have multiple observations ($N*T$) on each unit over time.

The convention is to refer to this data as either panel data or pooled time-series cross sectional data where:

Panel data is the term used when the data are dominated by large numbers of units (i) relative to time periods (t). The most often used panel data in political science is probably the 3 wave National Election Studies panel studies. For that $N \approx 2000$ and $T = 3$. We think of T as fixed and asymptotics hold as N approaches infinity.

Time series cross sectional data (tscs) is either dominated by time or simply has fewer units than the typical panel data set relative to time periods.

Examples here are countries over time where N is typically about 20 (as with EU analyses) and T may be between 10-30 years or more. Here we are interested in the differences between countries (they aren't just a sample, they are units we care about) – in panels we don't typically care about person *j* or *k*, their features are incidental). Here we typically think of N as fixed and asymptotics are in T.

In political science we typically work with tscs and this will be our focus.

Structure of the Data

The structure of the data is very important.

We need to preserve the time and space relationships in the data.

The structure is best thought of as stacking planes of data.

Each plane contains information on all variables and all time points for a specified unit or cross-section.

We stack the units over each other.

For example, pooling states over years.

We list all the time periods for state number 1 followed by all the time periods for state number 2 and so on.

[1] These notes come mostly from Suzanna De Boef's webpage for her course on PTSCS.

There are several ways the data can vary.

We may have a dependent variable that varies over time and between units; that is, it has within state variation and between state variation.

We may have an independent variable that varies over time but is the same for each point for all units. For example, national-level inflation when the units are states.

We may have an independent variable that does not vary over time but varies by units. For example, whether or not a state is in the South.

We may have an independent variable that varies across space and time just as the dependent variable does. For example, the % of a state's vote going to the Republican candidate in the previous election.

There are two types of formats that Stata recognizes.

This is what Stata calls "long format"

state	year	dep	indep1	indep2	indep3
1	1900	13.6	75	1	100
1	1901	13.9	82	1	110
.	.	.	.	.	.
.	.	.	.	.	.
2	1900	12.2	75	0	90
2	1901	12.1	82	0	95
.	.	.	.	.	.
.	.	.	.	.	.

You may find a data set where the data are stacked by time. Stata refers to this as "wide format" and can easily convert between the two formats.

Example of wide format:

state	year	dep	indep1	indep2	indep3
1	1900	13.6	75	1	100
2	1900	12.2	75	0	90
.	.	.	.	.	.
.	.	.	.	.	.
1	1901	13.9	82	1	110
2	1901	12.1	82	0	95
.	.	.	.	.	.
.	.	.	.	.	.

Advantages of TSCS Data

We may want to study both comparative and dynamic aspects.

Pooling allows us to use more information than we can gather for a purely cross sectional or time serial design.

However, it also adds to the level of difficulty of our modeling.

In particular, we need to worry about both time dependence and heteroskedasticity in the error term of a pooled model.

In fact, we have a 2-dimensional error term so that the possibility of violating error assumptions is higher than when modeling either cross sections or time series.

It is important to stress that there are many occasions when pooling is not a good strategy.

In particular, we would not want to pool units which are fundamentally different in a way that the underlying processes (the time series component) are unique and non-comparable.

We must be willing to argue that the independent variables have the same relationship or follow the same process in each country.

Deciding whether or not to pool should be the first stage of the analysis.

We ask ourselves first if the units are comparable and if the processes operating in the units is the same.

If this is not the case, rethink the decision to pool.

Some reasons that we might pool time series and cross sections.

1. Structure of the question.

Our focus may be on comparison – e.g. how are nations or state different.

From there we may want to know about similarities and differences in their dynamics.

If we are interested in understanding how attitudes toward democracy, for example, vary with current and past political and economic conditions as well as with demographic information, we need to compare people across contexts.

In a panel, we might examine repeated surveys of the same individuals in a single country over time as the political and economic context vary.

In tscs we might examine aggregate opinion of a random sample of individuals (a time series of opinion marginals) across countries as a function of time series of political and economic conditions and demographics.

In both cases, the units are different and there are dynamics in the differences; both space (individuals or countries) and time (politics, economic, etc.) interact.

2. Our ability to answer questions is often improved with tscs.

If a process is believed to unfold in a similar manner across different units, we can gain information about dynamics from the accumulation of time series for each cross section.

We may have too few time points to do useful time series or too few cases to do useful cross sectional analysis.

3. Pooling provides us with the kind of variation we need to answer questions we cannot answer with either time series or cross sectional data alone.

Single cross section: estimates reflect distinctions across individuals/subjects/nations – variation is across units so variables explain differences in units, which may not tell us what we want to know. If we can follow a nation's behavior over time, however, we'll be able to distinguish certain hypotheses about effects of some *x*, which may vary over time.

Single time series: if dynamics are important and autoregression strong, having lagged *x* on the right hand side will present a multicollinearity problem. If we have a set of panels then the different countries values of *x* will provide more information and reduce the problem.

4. Pooling allows us to deal with some types of measurement error and omitted variables problems associated with either the unit or the time series alone.

If we regress y_{it} on x_{it} and z_{it} and errors are iid over *i* and *t* with mean 0 and constant variance, we can use OLS and be fine: it will be consistent, unbiased and efficient.

But if we cannot measure z_{it} and if the z_{it} are correlated with x_{it}, then OLS will be biased.

How can pooled tscs help?

If $z_{it} = z_i$ for all *t* (That is, values of *z* are constant over time but vary over individuals/units), we can take the first difference of individual observations over time and obtain:

$$y_{it} - y_{it-1} = \beta'(x_{it} - x_{it-1}) + (\mu_{it} - \mu_{it-1}) + (z_{it} - z_{it-1})$$

Since the last term is zero, this eliminates the effect of *z* on *y* and our estimates of β'.

Second, if $z_{it} = z_t$ for all *i* (values of *z* are constant for each cross section at one point in time, but vary over time, we can take deviation from the mean across individuals at a give time and obtain:

$$y_{it} - \bar{y}_t = \beta'(x_{it} - \bar{x}_t) + (\mu_{it} - \bar{\mu}_t) + (z_{it} - \bar{z}_t)$$

Again, the last term is zero eliminating the effect of *z*.

Obviously, these two tricks require multiple observations on multiple cases.

OLS and Pooled Designs

Consider a simple pooled model:

$$y_{it} = \alpha + \beta' x_{it} + \mu_{it}$$

Where *y* is continuous, x_{it} is a vector of *k* exogenous variables, Ω (omega) is the *NT x TN* covariance matrix of errors with typical element, $E(\mu_{i,t}\mu_{j,s})$. The diagonal is composed of variances of the various *it*.

The off-diagonal elements give correlations of the errors across units and time:

$$\Omega = \begin{bmatrix} E(\mu_{11}^2) & E(\mu_{11}\mu_{12}) & \cdots & E(\mu_{11}\mu_{1T}) & E(\mu_{11}\mu_{21}) & E(\mu_{11}\mu_{22}) & \cdots & E(\mu_{11}\mu_{NT}) \\ E(\mu_{12}\mu_{11}) & E(\mu_{12}^2) & \cdots & E(\mu_{12}\mu_{1T}) & E(\mu_{12}\mu_{21}) & E(\mu_{12}\mu_{22}) & \cdots & E(\mu_{12}\mu_{NT}) \\ . & . & \cdots & . & . & . & \cdots & . \\ E(\mu_{1T}\mu_{11}) & E(\mu_{1T}\mu_{12}) & \cdots & E(\mu_{1T}^2) & E(\mu_{1T}\mu_{21}) & E(\mu_{1T}\mu_{22}) & \cdots & E(\mu_{1T}\mu_{NT}) \\ E(\mu_{21}\mu_{11}) & E(\mu_{21}\mu_{12}) & \cdots & E(\mu_{21}\mu_{1T}) & E(\mu_{21}^2) & E(\mu_{21}\mu_{22}) & \cdots & E(\mu_{21}\mu_{NT}) \\ E(\mu_{22}\mu_{11}) & E(\mu_{22}\mu_{12}) & \cdots & E(\mu_{22}\mu_{1T}) & E(\mu_{22}\mu_{21}) & E(\mu_{22}^2) & \cdots & E(\mu_{22}\mu_{NT}) \\ . & . & \cdots & . & . & . & \cdots & . \\ E(\mu_{NT}\mu_{11}) & E(\mu_{NT}\mu_{12}) & \cdots & E(\mu_{NT}\mu_{1T}) & E(\mu_{NT}\mu_{21}) & E(\mu_{NT}\mu_{22}) & \cdots & E(\mu_{NT}^2) \end{bmatrix}$$

The standard linear regression assumptions can then be written as:

1. $E(\mu_{it}) = 0$ for all *i* and all *t*.
2. $\text{var}(\mu_{it}) = \sigma^2$ for all *i* and all *t*. This is the homoskedasticity assumption.

3. $\text{cov}(\mu_{it}\mu_{jt}) = 0$ for any *i*, *j*, and *t*. Cross sectional independence; no unit behavior depends on the behavior of other units. The expected correlation between the errors for one unit at a given time period and another unit at the same time period is 0.

4. $\text{cov}(\mu_{it}\mu_{it-s}) = 0$ for any *I* and *t*. This is the no autocorrelation assumption.

5. $\mu_{it} \sim N(0, \sigma^2)$ Error is normally distributed.

If these assumptions hold, we can write the variance/covariance matrix as follows where *I* is the identity matrix.

$$\Omega = \begin{bmatrix} \sigma^2 I & 0 & \cdots & 0 \\ 0 & \sigma^2 I & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & \cdots & \sigma^2 I \end{bmatrix}$$

The equation as written also implies:

The effects, βs, are constant across units and over time, conditional on the *x*.

The intercepts, α, are constant across units and over time, conditional on the *x*.

In other words, all units must follow the same process or obey the same equation.

Potential Problems with OLS

OLS ignores the structure of the data; each case is treated as independent of all others.

The problem with pooled time series analysis is that both the traditional OLS assumptions and these last two assumptions may not be reasonable when we have panel or tscs data.

Another way to think of this is that the simple linear model we have specified implies homogenous intercepts and slopes and places a lot of restrictions on Ω.

These restrictions are often not valid in which case we need a different model and/or estimator.

Possible problems:

→ The assumption that slopes and intercepts are constant across units and over time is often untenable.

- if some states have systematically higher or lower values of the dependent variable than other states, given the independent variables, then we need to allow unique intercepts for each unit. Also, if the dynamic process that characterizes the relationships in the data change over time, we must allow for unique intercepts in these time periods.

- if relationships on some variables differ by unit (or over time) it will lead to bias.

→ Consider assumption 2, homoskedasticity. Typically some units will be more variable than others. This may occur for seemingly innocuous reasons such as sample size differences. For example, NY and California are bigger states with higher state incomes and heterogeneity across states will lead to unequal variances across the units, a problem that must be dealt with or estimates will be inefficient.

→ Assumption 3 is that the behavior of units is independent so that the errors across units are independent and there is no contemporaneous correlation in the errors. If the model omits a variable or there is a shock, which affects all units equally, then the errors across units will not be independent.

→ Assumption 4 assumes that there is no autocorrelation. This is almost always a problem and should make us suspicious about using OLS.

OLS and Heterogeneity

Begin thinking about modeling tscs in the simplest case when we satisfy typical regression assumptions, but are concerned about what is called heterogeneity bias.

Heterogeneity bias means that, if after controlling for x_{it} the units are still different or time effects are still different, then we have heterogeneity.

If we don't model it we will have biased estimates.

This should be obvious, because the model is not properly specified if the slopes and intercepts are not homogenous and we use the basic OLS model that assumes homogeneity.

Beginning with the basic OLS model:

$$y_{it} = \alpha + \beta' x_{it} + \mu_{it}$$

What if we have different intercepts for at least some of the cases?

That is, $\alpha_i \neq \alpha$.

We need to allow α to vary by unit.

What if $\beta_i \neq \beta$ and $\alpha_i \neq \alpha$?

We will need to estimate: $y_{it} = \alpha_i + \beta_i' x_{it} + \mu_{it}$

We can test hypotheses about the appropriateness of the homogeneity assumption using the analysis of covariance. From there, we will talk about ways to model heterogeneity in intercepts within this same framework.

Analysis of Covariance

If we have *N*T* observations on individuals over time, we want to use all available information to test hypotheses. In order to do this reliably, we need to be able to argue that the processes affecting all *i* and all *t* are the "same."

In particular, the slopes and intercepts need to be the same for each cross section and over time.

We can test these hypotheses using analysis of covariance.

Consider and unrestricted linear model where the slope and intercepts vary by *i* and by *t*:

$$y_{it} = \alpha_{it} + \beta_{it}' x_{it} + \mu_{it}$$

We can't estimate this equation (more parameters than degrees of freedom) so its purpose is descriptive. We want to know if we can restrict this equation to something we can estimate with pooled data.

We can do this by testing:

- whether slopes and intercepts are simultaneously homogenous among different individuals and different time periods.

- whether slopes are the same.

- whether intercepts are the same.

If we can't reject the null hypothesis that slopes and intercepts are homogenous, then we are good to pool and can use OLS. If we can reject the null hypothesis then we want to know whether the problem is in intercepts, slopes or both.

Here is what we do:

1. Assume a separate regression model for each unit, *i*:

$$y_{it} = \alpha_i + \beta_i x_{it} + \mu_{it}$$

This is the unrestricted equation.

We can restrict this model in 3 ways:

(a) slopes identical, intercepts vary, $H_1 : \beta_1 = \beta_2 = ... = \beta_n$

Giving us: $y_{it} = \alpha_i + \beta x_{it} + \mu_{it}$

(b) slopes and intercepts are the same by *I*, $H_2 : \beta_1 = \beta_2 = ... = \beta_n, \alpha_1 = \alpha_2 = ... = \alpha_n$

Giving us: $y_{it} = \alpha + \beta x_{it} + \mu_{it}$

(c) intercepts are same, slopes are different (we don't usually test this one since we tend to have different intercepts if we have different slopes).

$$y_{it} = \alpha_{it} + \beta_i x_{it} + \mu_{it}$$

Under the assumption that the errors are independently and normally distributed with variance σ_μ^2 we can estimate the unrestricted model and the models implied by the restrictions in (a) and (b) and then test whether the reduction in the residual sum of squares is significant, so that the restrictions are binding.

Call the RSS from the unrestricted model S_{un}

Call the RSS from the first restriction S_{slope}

and call the RSS from the restriction on slopes and intercepts, $S_{int,slope}$.

Then we can test H_1 as follows:

$$F_1 = \frac{(S_{slope} - S_{un})/[(N-1)k]}{S_{un}/[NT - N(k+1)]}$$

Where F_1 has degrees of freedom $(N-1)k$ and $NT - N(k+1)$.

The numerator degrees of freedom = the difference in the degrees of freedom of the slope-restricted and the unrestricted model, or the number of linear restrictions imposed by the restricted model.

In this case, that means $NT - N(k+1)$ (DF for unrestricted model) – $(N(T-1)-k)$ (DF for slope restricted model) = $NT - Nk - N - [NT - N - k] = k(N-1)$.

Note: The denominator degrees of freedom is equal to the degrees of freedom for the unrestricted model.

If F_1 is significant then the unrestricted equation is maintained – the restricted and unrestricted equations are significantly different in terms of the residual variance (the difference in what is explained by the unrestricted and restricted equations is significantly different, so we don't want to restrict the equation; we do significantly better without restrictions).

If not, then we can restrict the equation for homogenous intercepts.

To test H_2, that intercepts and slopes are homogenous across *i* we use the following *F* test:

$$F_2 = \frac{(S_{\text{int},slope} - S_{un})/[(N-1)(k+1)]}{S_{un}/[NT - N(k+1)]}$$

F_2 has degrees of freedom $(N-1)(k+1)$ and $NT - N(k+1)$.

Note: as when testing the first hypothesis, the numerator degrees of freedom is equal to the difference in the number of linear restrictions imposed by the restricted model relative to the unrestricted: *NT-N(k+1)* (DF for unrestricted model) – [NT-(k+1)] (DF for intercept and slope restricted model) = (N-1)(k+1). The denominator is the same as for F_1.

If we can reject H_2,, then the unrestricted equation is maintained.

If not, then we can restrict the equation for both homogenous slopes and homogenous intercepts and go ahead and pool.

We can also test the conditional hypothesis that the intercepts are homogenous given that the slopes are homogenous by comparing the RSS for the regression with only the slopes restricted with that where both the slopes and the intercepts are restricted.

$$F_3 = \frac{(S_{\text{int},slope} - S_{slope})/(N-1)}{S_{slope}/[N(T-1)-k]}$$

F_3 has degrees of freedom *(N-1)* and *N(T-1)-k*.

Note: The numerator DF is given by [NT-(k+1)]-(N(t-1)-k)=N-1, which is the difference in number of degrees of freedom from each model or the number of linear restrictions made by the restricted model relative to the unrestricted.

2. We should also test the same set of restrictions for *t*.

To do so, we assume a separate regression for each *t* and proceed as above.

$$y_{it} = \alpha_t + \beta_t x_{it} + \mu_{it}$$

The restrictions are similar:

(a) slopes are identical and intercepts vary: $H_1^{'} : \beta_1 = \beta_2 = ... = \beta_t$

$$y_{it} = \alpha_t + \beta x_{it} + \mu_{it}$$

With $F_1^{'} = \dfrac{(S_{slope} - S_{un})/[(T-1)k]}{S_{un}/[NT - T(k+1)]}$

Where F_1 has degrees of freedom *(T-1)k* and *NT-T(k+1)*.

(b) slopes and intercepts are the same by *t*, $H_2^{'} : \beta_1 = \beta_2 = ... = \beta_t, \alpha_1 = \alpha_2 = ... = \alpha_t$

$$y_{it} = \alpha + \beta x_{it} + \mu_{it}$$

With $F_2^{'} = \dfrac{(S_{slope} - S_{un})/[(T-1)(k+1)]}{S_{un}/[NT - T(k+1)]}$

Where $F_2^{'}$ has degrees of freedom *(T-1)(k+1)* and *NT-T(k+1).*

(c) intercepts are the same, slopes are different, $H_3^{'}$ (don't usually test this one, tend to have different intercepts if we have different slopes).

$$F_3^{'} = \frac{(S_{\text{int},slope} - S_{slope})/(T-1)}{S_{slope}/[T(N-1) - k]}$$

Where $F_3^{'}$ has degrees of freedom *(T-1)* and *T(N-1)*.

These test results tell us whether unconditional pooling is legitimate or may lead to biases and tells us a bit about how (if) we can approach modeling the data.

An Example

Data set on US House elections from 1956-1992. Used in De Boef and Stimson (1995) "The Dynamic Structure of Congressional Elections." *Journal of Politics* 57(3):630-648.

The data set includes state level measures of:

The Democratic percent of the two-party vote in a state.

State level democratic partisanship (Erikson, Wright, and McIver – varies across states and is constant over time).

National trends in Democratic partisanship – varies only over time, is constant across states.

Midterm drop off – coded 1 if a midterm under a Democratic president, -1 under a Republican president and 0 if not a midterm election.

Incumbency – measured as the percentage of Democratic candidates running in the state that are incumbents.

Party in the White House – if Democrats are in the White House, coded 1, else 0.

Stimson's public policy mood – larger values imply more liberal policy preferences.
(Deviations from a linear trend) in real disposable income (coded for the party in power).

Dummy variable for whether the state is in the South or not.

The dataset includes information on 18 elections in 47 states.

We can learn a lot just from some basic descriptive statistics in Stata.

We'll being looking at the percent of the 2-party vote that was cast for all Democrat candidates in a state's house delegation.

Look at the range and means of this variable by state using Stata's tabstat command:

```
tabstat dempct, statistics(mean sd min max) by(state) columns(statistics)f(%9.2f)
```

Results are:
Percent Democratic of Two Party Vote Share in the States: 1956-1992

State	Mean	SD	Min	Max	State	Mean	SD	Min	Max
CT -1788	52.39	5.24	38.94	62.27	ALABAMA	70.63	12.01	48.44	97.45
MAINE -1	45.96	9.85	26.87	61.54	ARKANSAS	58.66	22.14	20.99	100
MA -1788	64.86	9.76	48.93	82.71	FLORIDA	57.42	7.01	45.58	71.8
NH -1788	41.39	7.07	29.73	53.81	GEORGIA	82.21	11.46	61.4	100
RI -1790	59.23	10.5	35.88	75.71	MS -1817	76.26	18.03	51.57	100
VERMONT	30.07	14.85	0	51.47	NC -1789	55.7	6.48	47.13	69.08
DELAWARE	50.34	9.73	37.08	67.52	SC -1788	69.78	17.05	49.7	100
NJ -1787	52.53	4.82	43.53	64.82	TEXAS -1	69.82	11.6	54.07	92.24
NEW YORK	55.63	4.4	44.92	61.78	KENTUCKY	56.56	5.63	47.07	65.33
PA -1787	52.65	3.43	45.44	58.35	MARYLAND	59.51	5.49	51.26	68.06
ILLINOIS	52.95	3.89	47.28	61.74	OKLAHOMA	59.37	4.73	52.42	69.88
INDIANA	51.53	4.19	43.29	57.42	TENNESSE	58.01	7.58	40.63	74.86
MICHIGAN	53.6	3.76	47.21	60.44	WV -1863	63.16	7.53	53	76.79
OHIO -18	51.94	6.55	41.25	63.67	ARIZONA	44.87	7.14	29.58	53.15
WISCONSI	48.84	8.66	25.64	61.01	COLORADO	49.21	5.18	35.87	58.18
IOWA -18	49.88	3.38	45.6	57.39	IDAHO -1	45.94	5.98	36.55	58.19
KANSAS -	42.13	5.06	32.3	50.42	MONTANA	53	5.24	43.54	64.66
MINNESOT	53.66	4.03	47.73	59.86	NEVADA -	62.1	13.01	41.34	86.46

MISSOURI	59.38	4.79	50.03	69.68	NM -1912	51.81	6.42	40.5	61.49
NEBRASKA	37.85	7.43	22.4	48.64	UTAH -18	46.63	7.04	34.96	58.57
ND -1889	50.65	15.82	27.16	78.7	WYOMING	42.1	9.51	24.91	56.44
SD -1889	49.23	11.34	24.69	71.74	CA -1850	50.96	3.43	45.08	56.9
VIRGINIA	51.1	7.51	38.22	66.26	OREGON -	58.25	7.43	47.01	69.54
					WA -1889	53.31	6.52	38.34	65.95

This information is helpful as we think about the heterogeneity in the states. We notice that many of the states have similar mean values between 48 and 52% or so.

But many states are noticeably more Democratic (MA, RI, MO, MS, GA, FL, AR, AL, SC, TX, MD, OK, TN, WV, NV, OR) or Republican (NH, VT, KS, AZ, WY) suggesting that we must control for unit heterogeneity.

Many of the highly Democratic states are in the south (so that we might consider having a dummy variable for Southern states – as there already is).

The standard deviations vary from just over 3 to 22, suggesting heterogeneity as well, although controlling for incumbency may minimize that problem too.

Now that we've taken a peek at the data, we can test for the homogeneity of intercepts and/or slopes.

There are a few ways to do this in Stata.

The easiest of these is to use the ANOVA command.

Do-file example tests for homogeneity of intercepts and the slope of Democratic macropartisanship (which varies only over time).

Dealing with Unit Heterogeneity

Often the differences between units are not of real interest to the analyst.

Example: looking at a model of presidential election outcomes by states over time we are trying to assess the role of economic conditions on the vote.

The dependent variable is the Democrat percentage of the vote.

Some states will have a history of high levels of support for the Democratic candidates.

In this case, our interest is not in the structural differences in the states' average level of support but in how that level is affected by changes in the economy.

However, ignoring these differences between the states (using OLS) will result in bias in our estimates.

Alternatively, we might care about the unit effects – differences between countries, for example, might be of interest to us so that characterizing the heterogeneity is a goal.

Another way to think about this is that we cannot measure or don't have good data on the constructs that will explain country differences so that we have omitted variables.

As long as these omitted variables are constant across units (and/or over time) we can model the heterogeneity without introducing specification bias.

If fixed effects are needed and we leave them out, we have omitted variable bias that can be very serious if (a) the unit effects are non-trivial and (b) the unit variables are correlated with the *x*'s in the model.

There are two general ways to model this variation: Fixed effects (LSDV or standardizing the data) and Random Effects (GLS-error components) models.

Each makes different assumptions about the data and the process underlying the behavioral model.

Fixed Effects: Least Squares Dummy Variable Estimation (or the Covariance Model)

The simplest solution to heterogeneity in the units is to model it directly with separate intercepts.

This is the standard fixed effects model.

The assumption underlying this model is that each unit (or time period) has a unique but constant source of variation; that the variance or source of heterogeneity is fixed. (Here fixed does not mean non-stochastic. It means time invariant.)

Differences across units can be captured in differences in the constant.

We introduce dummies to model these unique sources of variation, to model the heterogeneity.

Thus, the fixed effects model is sometimes called the LSDV model or the variable-intercepts model (Hsaio).

Here we look only at unit effects but could consider time effects as well.

Because the fixed effects are either constant over time or across units, or both, they can be absorbed into the intercept and the regression will be unbiased and efficient.

The β will also be consistent as long as N or T or both tend to infinity, *even if these effects are correlated with the* x_{it}.

For fixed effects, the intercepts will be consistent only if *T* goes to infinity.

This latter point is very important and we will return to it later.

The general model for LSDV is:

$$y_{it} = \alpha_i + \tau_t + \beta' x_{it} + \mu_{it}$$

where $i=1,2,\ldots N$ cross sectional units and $t=1,2,\ldots T$ time points and $k=1,2,\ldots,K$ explanatory variables and where μ_{it} satisfies OLS assumptions.

Further, α_i give unit specific effects and τ_t give the time specific effects.

For now we will constrain the time effects to be zero and assume that our interest is in controlling for the unit-specific effects.

This is simply estimating a unique intercept for each unit and forcing the model through the origin – we may estimate an intercept and *n-1* dummies.

Each intercept reflects the variance unique to that cross-section.

In this way we can think of each intercept as the differences unique to each state that are not explained by the remaining independent variables in the model.

This has the effect of moving the regression line of each unit up or down while leaving all regression lines parallel.

We wish to model these differences so that the "true" effect of the x_{kit} can be estimated without bias.

These intercepts are each estimated with *t* observations so they won't be well estimated in sample that are dominated by N but have small T.

Reporting the coefficients may best be done with a graph – simplifies some a very long table.

We can test the hypothesis that all the constant terms are equal with an F test:

$$F = \frac{(R^2_{LSDV} - R^2_{Pooled})/(n-1)}{(1-R^2_{LSDV})/(nT-n-k)}$$

with $n-1$ and $nT-n-k$ degrees of freedom and where the pooled model is the restricted model with only one intercept.

Example: Democratic percent of the two-party vote in a state as a function of

State-level partisanship (Erikson, Wright and McIver – varies across states and is constant over time).

National trends in Democratic partisanship (varies only over time, is constant across states),

Midterm drop off (coded 1 if a midterm under a Democratic president, -1 under a Republican president, and 0 if not a midterm election).

Incumbency – measured as the percentage of Democratic candidates running in the state that are incumbents.

Party in the White House – if Democrats are in the White House, coded 1, else 0.

Stimson's public policy mood – larger values imply more liberal policy preferences.

(Deviations from a linear trend) in real disposable income (coded for the party in power).

Dummy variable for whether the state is in the South or not.

Start with OLS model:

```
. reg dempct d2p dmp midterm newinc south pip mood demrdi

      Source |       SS       df       MS              Number of obs =     846
-------------+------------------------------           F(  8,   837) =  114.37
       Model |  74212.8238     8  9276.60297           Prob > F      =  0.0000
    Residual |  67886.5185   837  81.1069516           R-squared     =  0.5223
-------------+------------------------------           Adj R-squared =  0.5177
       Total |  142099.342   845  168.164902           Root MSE      =  9.0059

------------------------------------------------------------------------------
      dempct |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         d2p |   .1709586   .0370731     4.61   0.000     .0981915    .2437258
         dmp |   .3037394   .0870339     3.49   0.001      .132909    .4745697
     midterm |  -2.189344   .7268242    -3.01   0.003    -3.615957   -.7627322
      newinc |   .5397455   .0256964    21.00   0.000     .4893086    .5901824
       south |   7.529627   .8699932     8.65   0.000     5.822003    9.237252
         pip |  -.5343645   .5169902    -1.03   0.302    -1.549114    .4803851
        mood |   .1500769   .0381021     3.94   0.000     .0752901    .2248637
      demrdi |   .1906955   .1801713     1.06   0.290    -.1629451    .5443361
       _cons |  -19.27943    7.68515    -2.51   0.012    -34.36386      -4.195
------------------------------------------------------------------------------
```

Now add state dummies and drop the variables that are constant across time within states – these are perfectly collinear with the state intercepts.

Create state dummies in Stata using xi command:

```
xi i.state
```

Then:

```
regress dempct  _Istate_2  _Istate_3 _Istate_4 _Istate_5 _Istate_6 _Istate_11
_Istate_12 _Istate_13 _Istate_14 _Istate_21 _Istate_22 _Istate_23 _Istate_24
_Istate_25 _Istate_31 _Istate_32 _Istate_33 _Istate_34 _Istate_35 _Istate_36
_Istate_37 _Istate_40 _Istate_41 _Istate_42 _Istate_43 _Istate_44 _Istate_46
_Istate_47 _Istate_48 _Istate_49 _Istate_51 _Istate_52  _Istate_53 _Istate_54
_Istate_56 _Istate_61 _Istate_62 _Istate_63 _Istate_64 _Istate_65 _Istate_66
_Istate_67 _Istate_68 _Istate_71 _Istate_72 _Istate_73  dmp midterm newinc pip
mood demrdi, noconstant

      Source |       SS       df       MS              Number of obs =     846
-------------+------------------------------           F( 52,   794) =  823.74
       Model |  2590334.79    52  49814.1306           Prob > F      =  0.0000
    Residual |  48015.5099   794  60.4729344           R-squared     =  0.9818
-------------+------------------------------           Adj R-squared =  0.9806
       Total |   2638350.3   846  3118.61738           Root MSE      =  7.7764
```

```
--------------------------------------------------------------------------------
      dempct |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+------------------------------------------------------------------
   _Istate_2 |  -1.901701   2.554018    -0.74   0.457    -6.915127    3.111724
   _Istate_3 |    9.71427   2.555035     3.80   0.000     4.698849    14.72969
       .
       .
       .
   Istate_71 |  -1.425734   2.543783    -0.56   0.575    -6.419069    3.567601
  _Istate_72 |   4.349738   2.548337     1.71   0.088    -.6525357    9.352013
  _Istate_73 |    .995723   2.543673     0.39   0.696    -3.997395    5.988842
         dmp |   .2595829   .0423569     6.13   0.000     .1764382    .3427275
     midterm |  -1.745385   .5964295    -2.93   0.004     -2.91615   -.5746199
      newinc |   .3964084   .0264054    15.01   0.000     .3445757    .4482411
         pip |  -.6320296   .4455296    -1.42   0.156    -1.506585    .2425255
        mood |    .140224     .02432     5.77   0.000     .0924849    .1879631
      demrdi |   .2530996   .1406088     1.80   0.072    -.0229094    .5291086
--------------------------------------------------------------------------------
```

Note that many of the dummy effects are edited out above to save space.

Or:

```
xi: regress  dempct _Istate_2 i.state dmp midterm newinc pip mood demrdi,
noconstant
```

An Alternative Fixed Effects Model: Demeaning Variables

We can remove the unit-specific effect from the data prior to estimation as well.

We can do this by scoring the dependent variable as deviations from the mean unit value (average for each state).

For example, if the mean value of the Democratic vote from president in Iowa is 52% we would recode each observed value based on its difference from this mean.

Thus, if 60% of voting Iowans supported Clinton in 1996, we could recode the support variable as 8 while a support level of 44% would have been recoded as -8.

This form of standardization would produce identical inferences as that achieved from a LSDV model.

The only observable differences occur when drawing interpretations from the model – you must remember to draw inferences with respect to the recoded support variable.

Further, the measures of model fit from the LSDV model will reflect the explained variation from the intercepts so will be substantially higher than in the standardized model.

Next I estimate the same model as above but I use Stata's `xtreg` command and the fixed effects (`fe`) option.

The estimation routine here removes the unit specific effects by demeaning the data before estimation:

$$(y_{it} - \bar{y}_i) = \beta'(x_{kit} - \bar{x}_{ki}) + (\mu_{it} - \bar{\mu}_i)$$

```
. xtreg dempct d2p dmp midterm newinc south pip mood demrdi, fe

Fixed-effects (within) regression               Number of obs      =       846
Group variable (i): state                       Number of groups   =        47

R-sq:  within  = 0.2740                         Obs per group: min =        18
       between = 0.7510                                        avg =      18.0
       overall = 0.4334                                        max =        18

                                                F(6,793)           =     49.88
corr(u_i, Xb)  = 0.4181                         Prob > F           =    0.0000

------------------------------------------------------------------------------
      dempct |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         d2p |  (dropped)
         dmp |   .3113361   .0751709     4.14   0.000     .1637787    .4588936
     midterm |   -1.91016   .6284517    -3.04   0.002    -3.143785    -.676534
      newinc |   .4005169   .0268666    14.91   0.000     .3477788     .453255
       south |  (dropped)
         pip |  -.6579351    .446698    -1.47   0.141    -1.534785    .2189152
        mood |   .1587114     .03292     4.82   0.000     .0940907    .2233321
      demrdi |   .1975975   .1556058     1.27   0.205    -.1078503    .5030454
       _cons |  -3.070453   6.548135    -0.47   0.639    -15.92418    9.783275
-------------+----------------------------------------------------------------
     sigma_u |    6.90082
     sigma_e |  7.7779304
         rho |  .44045884   (fraction of variance due to u_i)
------------------------------------------------------------------------------
F test that all u_i=0:     F(46, 793) =     7.16             Prob > F = 0.0000
```

The results here are identical to those in the last table, only the assessment of fit is different.

Stata does not assign the variance explained by the unit dummies to model fit.

The “Between” R^2 gives the correlation squared from the model where we use these estimates to predict the between effects model and the “overall” R^2 gives the correlation squared from the model where we use the estimates from the fixed effects model to predict the full model.

Note the variables that are dropped for perfect collinearity.

One advantage of standardizing the variables in this way is to get rid of the nuisance parameters (intercepts) to estimate.

In panels or tscs with large N this makes estimation easier and improves the consistency of the estimated effects.

Another way to do this is to use Stata’s `areg` command which filters the variables for you and gives just the β s.

The α s are “absorbed” or conditioned out of the model, hence the “a” in `areg`.

The `areg` command will also give an F test on the hypothesis that α s are simultaneously 0.

Here is the standardized estimation:

```
. areg dempct d2p dmp midterm newinc south pip mood demrdi, absorb(state)

                                                      Number of obs =     846
                                                      F(  6,   793) =   49.88
                                                      Prob > F      =  0.0000
                                                      R-squared     =  0.6624
                                                      Adj R-squared =  0.6403
                                                      Root MSE      =  7.7779

------------------------------------------------------------------------------
      dempct |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         d2p |  (dropped)
         dmp |   .3113361   .0751709     4.14   0.000     .1637787    .4588936
     midterm |   -1.91016   .6284517    -3.04   0.002    -3.143785    -.676534
      newinc |   .4005169   .0268666    14.91   0.000     .3477788     .453255
       south |  (dropped)
         pip |  -.6579351    .446698    -1.47   0.141    -1.534785    .2189152
        mood |   .1587114     .03292     4.82   0.000     .0940907    .2233321
      demrdi |   .1975975   .1556058     1.27   0.205    -.1078503    .5030454
       _cons |  -3.070453   6.548135    -0.47   0.639    -15.92418    9.783275
-------------+----------------------------------------------------------------
       state |        F(46, 793) =      7.156   0.000          (47 categories)
```

Notice that with the exception of fit, the results are identical to the `xtreg, fe` command.

Further, we can reject the null that the unit effects are unnecessary.

Note that from a statistical standpoint we could similarly constrain the unit effects to be zero and estimate unique time-specific effects for each time point using either LSDV or standardizing.

In this case standardizing the data would involve subtracting observed values from the mean value of the effects at one point in time across units.

More generally, it is possible to estimate fixed effects for both unit and time-specific heterogeneity.

However, in doing so we would use a lot of degrees of freedom and would likely attribute most of the variation to one of these types of effects leaving little variation to be explained by the independent variables in the model.

In these cases, the traditional regression extensions above are not very good and we need to consider alternatives to fixed effects OLS.

Random Effects: The Search for an Efficient Model

Fixed effects are unsatisfying in some respects.

1. If between-unit variation is large, controlling for this variation will mean that there is little left to be explained by the variables of interest. Even more problematic, if the explanatory variables are collinear with the unit effects, the estimates of (both) the explanatory variables (and the dummies in the LSDV formulation) will be inefficient.

2. We can lose many degrees of freedom if there are a large number of units. LSDV uses up to 1/T degrees of freedom, so this can be a big deal if you have small T.

3. Effects may be largely uniterpretable.

4. We may wish to draw inferences beyond the sample of units rather than measure unit differences.

5. We cannot simultaneously estimate a fixed effects regression and also include independent variables we care about that vary only by unit and not by time. These variables are perfectly collinear with the unit effects.

6. Similarly if the independent variables change only slowly, they will be highly collinear with the unit effects and their effects will be estimated poorly.

Given any or all of these concerns, we may consider a random effects model.

The way to think about this is that the unit effects are not explanations, but are measures of our ignorance (specific ignorance rather than general ignorance).

In this context, we can think of the unit effects as random variables (not fixed coefficients) whose distribution – both mean and variance – are informative, while the "effects" are not.

Rather than estimate unique intercepts for each unit, in the random effects model we estimate the parameters of the distributions of the difference pieces of error.

The estimated constant in the random effects model is the mean of this distribution of unobserved heterogeneity.

The specific (but unestimated) unit effects, the α_i, give the random heterogeneity specific to the i^{th} observation and is constant through time.

GLS or Error Components

We can rewrite the basic linear model breaking down the error into separate components resulting from 3 sources of variation: variation in the units, variation in time periods, and "true" random error variation.

Written this way the model is often called an "error components model":

$$y_{it} = \beta' x_{it} + \mu_{it}$$
$$and$$
$$\mu_{it} = \alpha_i + \tau_t + \psi_{it}$$

where *i=1,2,…,N* cross sectional units and *t=1,2,…,T* time points and *k=1,2,…,K* explanatory variables and where ψ_{it} (psi) satisfies OLS assumptions.

The α_i give the specific unit effects and τ_t give the time specific effects as before.

Consider the unit effects.

Think of the α_i as a random element that is specific to the unit.

It is like the random error, ψ_{it}, except that we have a single draw from the distribution that contributes to the error in each period.

Typically we assume that each error component is distributed normally, i.e. α_i i.i.d. $N(0,\sigma_\alpha^2)$.

Our goal is to separate the relative contributions of σ_α^2 and σ_ψ^2 (and $\sigma_{\tau_t}^2$ in the case of time effects) to the model.

This model is very similar to the fixed effects models but it makes different assumptions.

In particular, the random effects model makes two assumptions that are not necessary for the fixed effects models.

1. The unit effects (or time effects) are random draws from a common population.

2. Explanatory variables are strictly exogenous (error terms are uncorrelated with past, present, and future values of independent variables). This means that the unit effects must be uncorrelated with the x_{it}.

We can state the random effects model assumptions more formally as:

$$E(\alpha_i) = E(\tau_t) = E(\psi_{it}) = 0$$

$E(\alpha_i\tau_t) = E(\alpha_i\psi_{it}) = E(\tau_t\psi_{it}) = 0$ (the unit effects and true random error must be independent, same for time effects and random error, same for unit and time effects).

$E(\alpha_i\alpha_j) = \sigma_\alpha^2$ if $i=j$, 0 otherwise.

$E(\tau_t\tau_s) = \sigma_\tau^2$ if $t=s$, 0 otherwise.

$E(\psi_{it}\psi_{js}) = \sigma_\psi^2$ if $i=j$, $t=s$, 0 otherwise.

$$E(\alpha_i x_{it}) = E(\tau_t x_{it}) = E(\psi_{it} x_{it}) = 0$$

When these assumptions hold, then the variances of y_{it} conditional on the x_{it} is exactly equal to the sum of the variances of the units, the time periods and the random error variance,

$$= \sigma_\alpha^2 + \sigma_\tau^2 + \sigma_\psi^2$$

To do GLS we use the inverse square root of the Ω matrix which we use to transform each observation.

We use: $\theta = 1 - \frac{\sigma_\psi}{\sqrt{T\sigma_\alpha^2 + \sigma_\psi^2}}$

For the dependent variable we transform the data: $y_{it}^* = y_{it} - \theta \overline{y}_{it}$

And for each independent variable: $x_{it}^* = x_{it} - \theta \overline{x}_{it}$

We can see the relationship between the random effects and fixed effects model by noting that if $\theta = 1$ the results of the two are identical.

GLS transforms the OLS model with a general covariance matrix to another linear equation where the error is well behaved.

The problem is that we don't know the true variance/covariance matrix and must use an estimate – this is referred to as feasible GLS or FGLS.

In state, the random effects model is the default.

It is a matrix weighted average of the estimates produced by the between and within (fixed effects) estimators.

In other words, it weights each observation from the fixed effects model by θ rather than subtracting the unit means, we subtract some fraction of the unit means:

$$y_{it} - \theta \overline{y}_{it} = (1-\theta)v + \beta'(x_{it} - \theta \overline{x}_i) + [(1-\theta)\alpha_i + (\mu - it - \theta \overline{\mu}_i)]$$

To estimate the random effects model in Stata:

```
xtreg dempct d2p dmp midterm newinc south pip mood demrdi
```

The results are very similar to the results of the fixed effects estimation.

The biggest difference is that we now have estimates of the effects of state partisanship and being a southern state, both of which were perfectly collinear with the unit dummies and dropped from the fixed effects analysis.

The overall R^2 is slightly larger also.