# Panel data methods for microeconometrics using Stata

A. Colin Cameron
Univ. of California - Davis

Prepared for West Coast Stata Users' Group Meeting
Based on A. Colin Cameron and Pravin K. Trivedi,
*Microeconometrics using Stata, Stata Press, forthcoming.*

October 25, 2007

# 1. Introduction

**Panel data** are repeated measures on individuals $(i)$ over time $(t)$.

**Regress** $y_{it}$ on $\mathbf{x}_{it}$ for $i = 1, ..., N$ and $t = 1, ..., T$.

**Complications** compared to cross-section data:

1. **Inference:** correct (inflate) standard errors.
   This is because each additional year of data is not independent of previous years.
2. **Modelling:** richer models and estimation methods are possible with repeated measures.
   Fixed effects and dynamic models are examples.
3. **Methodology:** different areas of applied statistics may apply different methods to the same panel data set.

This talk: **overview** of panel data methods and `xt` commands for **Stata 10** most commonly used by **microeconometricians**.

Three **specializations** to general panel methods:

1. **Short panel:** data on many individual units and few time periods. Then data viewed as clustered on the individual unit. Many panel methods also apply to clustered data such as cross-section individual-level surveys clustered at the village level.
2. **Causation from observational data**: use repeated measures to estimate key marginal effects that are causative rather than mere correlation. Fixed effects: assume time-invariant individual-specific effects. IV: use data from other periods as instruments.
3. **Dynamic models:** regressors include lagged dependent variables.

# Outline

# 2.1 Some basic considerations

1. **Regular time intervals** assumed.
2. **Unbalanced** panel okay (xt commands handle unbalanced data). [Should then rule out selection/attrition bias].
3. **Short panel** assumed, with $T$ small and $N \rightarrow \infty$. [Versus long panels, with $T \rightarrow \infty$ and $N$ small or $N \rightarrow \infty$.]
4. **Errors are correlated.** [For short panel: panel over $t$ for given $i$, but not over $i$.]
5. **Parameters** may vary over individuals or time. **Intercept:** Individual-specific effects model (fixed or random effects). **Slopes:** Pooling and random coefficients models.
6. **Regressors**: time-invariant, individual-invariant, or vary over both.
7. **Prediction**: ignored. [Not always possible even if marginal effects computed.]
8. **Dynamic models**: possible. [Usually static models are estimated.]

# 2.2 Basic linear panel models

- **Pooled model (or population-averaged)**

$$y_{it} = \alpha + \mathbf{x}_{it}'\boldsymbol{\beta} + u_{it}. \tag{1}$$

- **Two-way effects model** allows intercept to vary over $i$ and $t$

$$y_{it} = \alpha_i + \gamma_t + \mathbf{x}_{it}'\boldsymbol{\beta} + \varepsilon_{it}. \tag{2}$$

- **Individual-specific effects model**

$$y_{it} = \alpha_i + \mathbf{x}_{it}'\boldsymbol{\beta} + \varepsilon_{it}, \tag{3}$$

  for short panels where time-effects are included as dummies in $\mathbf{x}_{it}$.

- **Random coefficients model** allows slopes to vary over $i$

$$y_{it} = \alpha_i + \mathbf{x}_{it}'\boldsymbol{\beta}_i + \varepsilon_{it}. \tag{4}$$

# 2.2 Fixed effects versus random effects

- Individual-specific effects model: $y_{it} = \mathbf{x}'_{it}\boldsymbol{\beta} + (\alpha_i + \varepsilon_{it})$.
- **Fixed effects (FE):**
  - $\alpha_i$ is possibly correlated with $\mathbf{x}_{it}$
  - regressor $\mathbf{x}_{it}$ can be **endogenous**
    (though only wrt a time-invariant component of the error)
  - can consistently estimate $\boldsymbol{\beta}$ for time-varying $\mathbf{x}_{it}$
    (mean-differencing or first-differencing eliminates $\alpha_i$)
  - cannot consistently estimate $\alpha_i$ if short panel
  - prediction is not possible
  - $\boldsymbol{\beta} = \partial \mathrm{E}[y_{it}|\alpha_i, \mathbf{x}_{it}]/\partial \mathbf{x}_{it}$

- **Random effects (RE) or population-averaged (PA)**
    - $\alpha_i$ is purely random (usually iid $(0, \sigma_\alpha^2)$).
    - regressor $\mathbf{x}_{it}$ must be **exogenous**
    - corrects standard errors for equicorrelated clustered errors
    - prediction is possible
    - $\boldsymbol{\beta} = \partial \mathrm{E}[y_{it}|\mathbf{x}_{it}]/\partial \mathbf{x}_{it}$
- **Fundamental divide**
    - Microeconometricians: fixed effects
    - Many others: random effects.

# 2.3 Robust inference

- Many methods assume $\varepsilon_{it}$ and $\alpha_i$ (if present) are iid.
- Yields **wrong standard errors** if heteroskedasticity or if errors not equicorrelated over time for a given individual.
- For short panel can relax and use **cluster-robust inference**.
  - Allows heteroskedasticity and general correlation over time for given $i$.
  - Independence over $i$ is still assumed.
- Use option `vce(cluster)` if available (`xtreg`, `xtgee`).
- This is not available for many `xt` commands.
  - then use option `vce(boot)` or `vce(cluster)`
  - but only if the estimator being used is still consistent.

# 2.4 Stata linear panel commands

| | |
|---|---|
| **Panel summary** | `xtset; xtdescribe; xtsum; xtdata;` `xtline; xttab; xttran` |
| **Pooled OLS** | `regress` |
| **Feasible GLS** | `xtgee, family(gaussian)` `xtgls; xtpcse` |
| **Random effects** | `xtreg, re; xtregar, re` |
| **Fixed effects** | `xtreg, fe; xtregar, fe` |
| **Random slopes** | `xtmixed; quadchk; xtrc` |
| **First differences** | `regress` (with differenced data) |
| **Static IV** | `xtivreg; xthtaylor` |
| **Dynamic IV** | `xtabond; xtdpdsys; xtdpd` |

# 3.1 Example: wages

- PSID wage data 1976-82 on 595 individuals. Balanced.
- Source: Baltagi and Khanti-Akom (1990).
  [Corrected version of Cornwell and Rupert (1998).]
- Goal: estimate causative effect of education on wages.
- Complication: education is time-invariant in these data.
  Rules out fixed effects.
  Need to use IV methods (Hausman-Taylor).

# 3.2 Reading in panel data

- `xt` commands require data to be in **long form.**
  Then each observation is an individual-time pair.
- Original data are often in **wide form.**
  Then an observation combines all time periods for an individual, or all individuals for a time period.
- Use `reshape long` to convert from wide to long.
- `xtset` is used to define *i* and *t*.
  - `xtset id t` is an example
  - allows use of panel commands and some time series operators.

# 3.3 Summarizing panel data

```
. summarize
```

| Variable | Obs | Mean | Std. Dev. | Min | Max |
|---|---|---|---|---|---|
| exp | 4165 | 19.85378 | 10.96637 | 1 | 51 |
| wks | 4165 | 46.81152 | 5.129098 | 5 | 52 |
| occ | 4165 | .5111645 | .4999354 | 0 | 1 |
| ind | 4165 | .3954382 | .4890033 | 0 | 1 |
| south | 4165 | .2902761 | .4539442 | 0 | 1 |
| smsa | 4165 | .6537815 | .475821 | 0 | 1 |
| ms | 4165 | .8144058 | .3888256 | 0 | 1 |
| fem | 4165 | .112605 | .3161473 | 0 | 1 |
| union | 4165 | .3639856 | .4812023 | 0 | 1 |
| ed | 4165 | 12.84538 | 2.787995 | 4 | 17 |
| blk | 4165 | .0722689 | .2589637 | 0 | 1 |
| lwage | 4165 | 6.676346 | .4615122 | 4.60517 | 8.537 |
| id | 4165 | 298 | 171.7821 | 1 | 595 |
| t | 4165 | 4 | 2.00024 | 1 | 7 |

- **describe, summarize and tabulate** confound cross-section and time series variation.
- Instead use **specialized panel commands**:
    - `xtdescribe`: extent to which panel is unbalanced
    - `xtsum`: separate within (over time) and between (over individuals) variation
    - `xttab`: tabulations within and between for discrete data e.g. binary
    - `xttrans`: transition frequencies for discrete data
    - `xtline`: time series plot for each individual on one chart
    - `xtdata`: scatterplots for within and between variation.

1. Pooled OLS: OLS of $y_{it}$ on $\mathbf{x}_{it}$.
2. Between estimator: OLS of $\bar{y}_i$ on $\bar{\mathbf{x}}_i$.
3. Random effects estimator: FGLS in RE model.
   Equals OLS of $(y_{it} - \widehat{\theta}_i \bar{y}_i)$ on $(\mathbf{x}_{it} - \widehat{\theta}_i \bar{\mathbf{x}}_i)$;
   $\theta_i = 1 - \sqrt{\sigma_\varepsilon^2 / (T_i \sigma_\alpha^2 + \sigma_\varepsilon^2)}$.
4. Within estimator or FE estimator: OLS of $(y_{it} - \bar{y}_i)$ on $(\mathbf{x}_{it} - \bar{\mathbf{x}}_i)$.
5. First difference estimator: OLS of $(y_{it} - y_{i,t-1})$ on $(\mathbf{x}_{it} - \mathbf{x}_{i,t-1})$.

- Implementation:
  - `xtreg` does 2-4 with options `be`, `fe`, `re`
  - `xtgee` does 3 (with option `exchangeable`)
  - `regress` does 1 and 5.
- Only 4. and 5. give consistent estimates of $\boldsymbol{\beta}$ in FE model.

# 4.2 Example

| Variable | ols | olsrob | fe | ferob | re | rerob |
|---|---|---|---|---|---|---|
| exp | 0.0447 | 0.0447 | 0.1138 | 0.1138 | 0.0889 | 0.0889 |
| | 0.0024 | 0.0054 | 0.0025 | 0.0044 | 0.0028 | 0.0040 |
| exp2 | -0.0007 | -0.0007 | -0.0004 | -0.0004 | -0.0008 | -0.0008 |
| | 0.0001 | 0.0001 | 0.0001 | 0.0001 | 0.0001 | 0.0001 |
| wks | 0.0058 | 0.0058 | 0.0008 | 0.0008 | 0.0010 | 0.0010 |
| | 0.0012 | 0.0019 | 0.0006 | 0.0009 | 0.0007 | 0.0009 |
| ed | 0.0760 | 0.0760 | 0.0000 | 0.0000 | 0.1117 | 0.1117 |
| | 0.0022 | 0.0052 | 0.0000 | 0.0000 | 0.0061 | 0.0084 |
| _cons | 4.9080 | 4.9080 | 4.5964 | 4.5964 | 3.8294 | 3.8294 |
| | 0.0673 | 0.1400 | 0.0389 | 0.0649 | 0.0936 | 0.1334 |
| N | 4165.0000 | 4165.0000 | 4165.0000 | 4165.0000 | 4165.0000 | 4165.0000 |
| r2 | 0.2836 | 0.2836 | 0.6566 | 0.6566 | | |
| r2_o | | | 0.0476 | 0.0476 | 0.1830 | 0.1830 |
| r2_b | | | 0.0276 | 0.0276 | 0.1716 | 0.1716 |
| r2_w | | | 0.6566 | 0.6566 | 0.6340 | 0.6340 |
| sigma_u | | | 1.0362 | 1.0362 | 0.3195 | 0.3195 |
| sigma_e | | | 0.1522 | 0.1522 | 0.1522 | 0.1522 |
| rho | | | 0.9789 | 0.9789 | 0.8151 | 0.8151 |

legend: b/se

- Coefficients vary considerably across OLS, FE and RE estimators.
- Cluster-robust standard errors (suffix rob) larger even for FE and RE.
- Coefficient of `ed` not identified for FE as time-invariant regressor.

# 4.3 Fixed effects versus random effects

- Use **Hausman test** to discriminate between FE and RE.
  - If fixed effects: FE consistent and RE inconsistent.
  - If not fixed effects: FE consistent and RE consistent.
  - So see whether difference between FE and RE is zero.

$$\mathrm{H} = \left(\widetilde{\beta}_{1,\mathrm{RE}} - \widehat{\beta}_{1,\mathrm{FE}}\right)' \left[\widehat{\mathrm{Cov}}[\widetilde{\beta}_{1,\mathrm{RE}} - \widehat{\beta}_{1,\mathrm{FE}}]\right]^{-1} \left(\widetilde{\beta}_{1,\mathrm{RE}} - \widehat{\beta}_{1,\mathrm{W}}\right),$$

  where $\beta_1$ corresponds to time-varying regressors (or a subset of these).

- Problem: `hausman` command assumes RE is fully efficient.
  But not the case here as robust se's for RE differ from default se's.
  So `hausman` is **incorrect**.
- Instead implement Hausman test using `suest` or panel bootstrap or Wooldridge (2002) robust version of Hausman test.

- Consider model with possibly transformed variables:

$$y^*_{it} = \alpha + \mathbf{x}^{*\prime}_{it}\boldsymbol{\beta} + u_{it},$$

where $y^*_{it} = y_{it}$ or $y^*_{it} = \bar{y}_i$ for BE or $y^*_{it} = (y_{it} - \bar{y}_i)$ for FE or $y^*_{it} = (y_{it} - \theta_i \bar{y}_i)$ for RE.

- OLS is **inconsistent** if $\mathrm{E}[u_{it}|\mathbf{x}^*_{it}] = 0$.
- So do **IV estimation** with **instruments** $\mathbf{z}^*_{it}$ satisfy $\mathrm{E}[u_{it}|\mathbf{z}^*_{it}] = 0$.
- Command `xtivreg` is used, with options `be`, `re` or `fe`.
- This command does not have option for robust standard errors.

# 5.2 Hausman-Taylor IV estimator

- Problem in the fixed effects model
  - If an **endogenous** regressor is time-invariant
  - Then FE estimator **cannot identify** $\boldsymbol{\beta}$ (as time-invariant).
- Solution:
  - Assume the endogenous regressor is **correlated only with** $\alpha_i$ (and not with $\varepsilon_{it}$)
  - Use exogenous time-varying regressors $\mathbf{x}_{it}$ from other periods as instruments
- Command `xthtaylor` does this (and has option `amacurdy`).

# 6.1 Linear dynamic panel models

- Simple dynamic model regresses $y_{it}$ in **polynomial in time**.
  - e.g. Growth curve of child height or IQ as grow older
  - use previous models with $\mathbf{x}_{it}$ polynomial in time or age.

- Richer dynamic model regresses $y_{it}$ on **lags** of $y_{it}$.

# 6.2 Linear dynamic panel models with individual effects

- **Leading example**: AR(1) model with individual specific effects

$$y_{it} = \gamma y_{i,t-1} + \mathbf{x}_{it}'\boldsymbol{\beta} + \alpha_i + \varepsilon_{it}.$$

- Three reasons for $y_{it}$ being serially correlated over time:
  - **True state dependence:** via $y_{i,t-1}$
  - **Observed heterogeneity:** via $\mathbf{x}_{it}$ which may be serially correlated
  - **Unobserved heterogeneity:** via $\alpha_i$
- Focus on case where $\alpha_i$ is a **fixed effect**
  - FE estimator is now inconsistent (if short panel)
  - Instead use Arellano-Bond estimator

- **First-difference** to eliminate $\alpha_i$ (rather than mean-difference)

$$(y_{it} - y_{i,t-1}) = \gamma(y_{i,t-1} - y_{i,t-2}) + (\mathbf{x}_{it} - \mathbf{x}'_{i,t-1})\boldsymbol{\beta} + (\varepsilon_{it} - \varepsilon_{i,t-1}).$$

- **OLS inconsistent** as $(y_{i,t-1} - y_{i,t-2})$ correlated with $(\varepsilon_{it} - \varepsilon_{i,t-1})$ (even under assumption $\varepsilon_{it}$ is serially uncorrelated).
- But $y_{i,t-2}$ is not correlated with $(\varepsilon_{it} - \varepsilon_{i,t-1})$,
  so can use $y_{i,t-2}$ as an **instrument** for $(y_{i,t-1} - y_{i,t-2})$.
- Arellano-Bond is a variation that uses **unbalanced set** of instruments with further lags as instruments.
  For $t = 3$ can use $y_{i1}$, for $t = 4$ can use $y_{i1}$ and $y_{i2}$, and so on.
- Stata commands
  - `xtabond` for Arellano-Bond
  - `xtdpdsys` for Blundell-Bond (more efficient than `xtabond`)
  - `xtdpd` for more complicated models than `xtabond` and `xtdpdsys`.

# 7.1 Long panels

- For **short panels** asymptotics are $T$ fixed and $N \rightarrow \infty$.
- For **long panels** asymptotics are for $T \rightarrow \infty$
  - A dynamic model for the errors is specified, such as AR(1) error
  - Errors may be correlated over individuals
  - Individual-specific effects can be just individual dummies
  - Furthermore if $N$ is small and $T$ large can allow slopes to differ across individuals and test for poolability.

# 7.2 Commands for long panels

- Models with **stationary errors:**
  - `xtgls` allows several different models for the error
  - `xtpcse` is a variation of `xtgls`
  - `xtregar` does FE and RE with AR(1) error
- Models with **nonstationary errors** (currently active area):
  - As yet no Stata commands
  - Add-on `levinlin` does Levin-Lin-Chu (2002) panel unit root test
  - Add-on `ipshin` does Im-Pesaran-Shin (1997) panel unit root test in heterogeneous panels
  - Add-on `xtpmg` for does Pesaran-Smith and Pesaran-Shin-Smith estimation for nonstationary heterogeneous panels with both $N$ and $T$ large.

# 8.1 Random coefficients model

- Generalize random effects model to **random slopes**.
- Command xtrc estimates the **random coefficients model**

$$y_{it} = \alpha_i + \mathbf{x}'_{it}\boldsymbol{\beta}_i + \varepsilon_{it},$$

where $(\alpha_i, \boldsymbol{\beta}_i)$ are iid with mean $(\alpha, \boldsymbol{\beta})$ and variance matrix $\Sigma$ and $\varepsilon_{it}$ is iid.

- No vce(robust) option but can use vce(boot) if short panel.

# 8.2 Mixed or multi-level or hierarchical model

- Not used in microeconometrics but used in many other disciplines.
- Stack all observations for individual $i$ and specify

$$\mathbf{y}_i = \mathbf{X}_i \boldsymbol{\beta} + \mathbf{Z}_i \mathbf{u}_i + \boldsymbol{\varepsilon}_i$$

  where $\mathbf{u}_i$ is iid $(\mathbf{0}, \mathbf{G})$ and $\mathbf{Z}_i$ is called a design matrix.
- Random effects: $\mathbf{Z}_i = \mathbf{e}$ (a vector of ones) and $\mathbf{u}_i = \alpha_i$
- Random coefficients: $\mathbf{Z}_i = \mathbf{X}_i$.
- Other models including **multi-level models** are possible.
- Command `xtmixed` estimates this model.

# 9.1 Clustered data

- Consider data on individual $i$ in village $j$ with **clustering on village**.
- A **cluster-specific model** (here village-specific) specifies

$$y_{ji} = \alpha_i + \mathbf{x}_{ji}'\boldsymbol{\beta} + \varepsilon_{ji}.$$

- Here clustering is on village (not individual) and the repeated measures are over individuals (not time).
- Use `xtset village id`
- Assuming **equicorrelated errors** can be more reasonable here than with panel data (where correlation dampens over time).
  So perhaps less need for `vce(cluster)` after `xtreg`

# 9.2 Estimators for clustered data

- If $\alpha_i$ is **random** use:
  - `regress` with option `vce(cluster village)`
  - `xtreg,re`
  - `xtgee` with option `exchangeable`
  - `xtmixed` for richer models of error structure
- If $\alpha_i$ is **fixed** use:
  - `xtreg,fe`

# 10.1 Nonlinear panel models overview

- **General approaches** similar to linear case
  - Pooled estimation or population-averaged
  - Random effects
  - Fixed effects
- **Complications**
  - Random effects often not tractable so need numerical integration
  - Fixed effects models in short panels are generally not estimable due to the incidental parameters problem.
- Here we consider **short panels** throughout.
- **Standard nonlinear models** are:
  - Binary: logit and probit
  - Counts: Poisson and negative binomial
  - Truncated: Tobit

# 10.2 Nonlinear panel models

- A **pooled** or **population-averaged model** may be used.
This is same model as in cross-section case, with adjustment for correlation over time for a given individual.
- A **fully parametric model** may be specified, with conditional density

$$f(y_{it}|\alpha_i, \mathbf{x}_{it}) = f(y_{it}, \alpha_i + \mathbf{x}_{it}'\boldsymbol{\beta}, \gamma), \quad t = 1, ..., T_i, \ i = 1, ...., N, \quad (5)$$

where $\gamma$ denotes additional model parameters such as variance parameters and $\alpha_i$ is an individual effect.
- A **conditional mean model** may be specified, with **additive effects**

$$\mathrm{E}[y_{it}|\alpha_i, \mathbf{x}_{it}] = \alpha_i + g(\mathbf{x}_{it}'\boldsymbol{\beta}) \quad (6)$$

or **multiplicative effects**

$$\mathrm{E}[y_{it}|\alpha_i, \mathbf{x}_{it}] = \alpha_i \times g(\mathbf{x}_{it}'\boldsymbol{\beta}). \quad (7)$$

# 10.3 Nonlinear panel commands

| | Counts | Binary |
|---|---|---|
| **Pooled** | poisson | logit |
| | negbin | probit |
| **GEE (PA)** | xtgee,family(poisson) | xtgee,family(binomial) link(logit |
| | xtgee,family(nbinomial) | xtgee,family(poisson) link(probit |
| **RE** | xtpoisson, re | xtlogit, re |
| | xtnegbin, fe | xtprobit, re |
| **Random slopes** | xtmepoisson | xtmelogit |
| **FE** | xtpoisson, fe | xtlogit, fe |
| | xtnegbin, fe | |

plus `tobit` and `xttobit`.

# 11.1 Pooled or Population-averaged estimation

- Extend pooled OLS
  - Give the usual cross-section command for conditional mean models or conditional density models but then get cluster-robust standard errors
  - Probit example:

```
probit y x, vce(cluster id)
```

or

```
xtgee y x, fam(binomial) link(probit) corr(ind)
vce(cluster id)
```

- Extend pooled feasible GLS
  - Estimate with an assumed correlation structure over time
  - Equicorrelated probit example:

```
xtprobit y x, pa vce(boot)
```

or

```
xtgee y x, fam(binomial) link(probit) corr(exch)
vce(cluster id)
```

# 11.2 Random effects estimation

- Assume individual-specific effect $\alpha_i$ has specified distribution $g(\alpha_i|\boldsymbol{\eta})$.
- Then the unconditional density for the $i^{th}$ observation is

$$
\begin{aligned}
& f(y_{it},\ldots,y_{iT}|\mathbf{x}_{i1},\ldots,\mathbf{x}_{iT},\boldsymbol{\beta},\gamma,\boldsymbol{\eta}) \\
= & \int \left[\prod_{t=1}^{T} f(y_{it}|\mathbf{x}_{it},\alpha_i,\boldsymbol{\beta},\gamma)\right] g(\alpha_i|\boldsymbol{\eta})d\alpha_i. \qquad (8)
\end{aligned}
$$

- **Analytical solution:**
  - For Poisson with gamma random effect
  - For negative binomial with gamma effect
  - Use `xtpoisson, re` and `xtnbreg, re`
- **No analytical solution:**
  - For other models.
  - Instead use numerical integration (only univariate integration is required).
  - Assume normally distributed random effects.
  - Use `re` option for `xtlogit`, `xtprobit`
  - Use `normal` option for `xtpoisson` and `xtnegbin`

# 11.2 Random slopes estimation

- Can extend to **random slopes**.
  - Nonlinear generalization of `xtmixed`
  - Then higher-dimensional numerical integral.
  - Use adaptive Gaussian quadrature
- Stata commands are:
  - `xtmelogit` for binary data
  - `xtmepoisson` for counts
- Stata add-on that is very rich:
  - `gllamm` (generalized linear and latent mixed models)
  - Developed by Sophia Rabe-Hesketh and Anders Skrondal.

# 11.3 Fixed effects estimation

- In general not possible in short panels.
- **Incidental parameters problem:**
  - $N$ fixed effects $\alpha_i$ plus $K$ regressors means $(N + K)$ parameters
  - But $(N + K) \to \infty$ as $N \to \infty$
  - Need to eliminate $\alpha_i$ by some sort of differencing
  - possible for Poisson, negative binomial and logit.
- Stata commands
  - `xtlogit, fe`
  - `xtpoisson, fe` (better to use `xtpqml` as robust se's)
  - `xtnegbin, fe`
- Fixed effects extended to **dynamic models** for logit and probit. No Stata command.

# 12. Conclusion

- Stata provides commands for panel models and estimators commonly used in microeconometrics and biostatistics.
- Stata also provides diagnostics and postestimation commands, not presented here.
- The emphasis is on short panels. Some commands provide cluster-robust standard errors, some do not.
- A big distinction is between fixed effects models, emphasized by microeconometricians, and random effects and mixed models favored by many others.
- Extensions to nonlinear panel models exist, though FE models may not be estimable with short panels.
- This presentation draws on two chapters in Cameron and Trivedi, *Microeconometrics using Stata*, forthcoming.

# Book Outline

For Cameron and Trivedi, *Microeconometrics using Stata*, forthcoming.

1. Stata basics
2. Data management and graphics
3. Linear regression basics
4. Simulation
5. GLS regression
6. Linear instrumental variable regression
7. Quantile regression
8. Linear panel models
9. Nonlinear regression methods
10. Nonlinear optimization methods
11. Testing methods
12. Bootstrap methods

# Book Outline (continued)

13. Binary outcome models
14. Multinomial models
15. Tobit and selection models
16. Count models
17. Nonlinear panel models
18. Topics

A. Programming in Stata

B. Mata

# Econometrics graduate-level panel data texts

- Comprehensive panel texts
  - Baltagi, B.H. (1995, 2001, 200?), *Econometric Analysis of Panel Data*, 1st and 2nd editions, New York, John Wiley.
  - Hsiao, C. (1986, 2003), *Analysis of Panel Data*, 1st and 2nd editions, Cambridge, UK, Cambridge University Press.
- More selective advanced panel texts
  - Arellano, M. (2003), *Panel Data Econometrics*, Oxford, Oxford University Press.
  - Lee, M.-J. (2002), *Panel Data Econometrics: Methods-of-Moments and Limited Dependent Variables*, San Diego, Academic Press.
- Texts with several chapters on panel
  - Cameron, A.C. and P.K. Trivedi (2005), *Microeconometrics: Methods and Applications*, New York, Cambridge University Press.
  - Greene, W.H. (2003), *Econometric Analysis*, fifth edition, Upper Saddle River, NJ, Prentice-Hall.
  - Wooldridge, J.M. (2002, 200?), *Econometric Analysis of Cross Section and Panel Data*, Cambridge, MA, MIT Press.