Handout 2: Advanced Panel Data Models

Master in Data Science for Decision Making
Barcelona School of Economics

Laura Mayoral

IAE and BSE

Barcelona, Winter 2025

Introduction

- Previous handout: linear and static panel data models

- These models are often not suitable.

- For instance, when considering binary data: (y_{it} takes values 0 and 1):

  - Effect of an information campaign on smoking behavior (1 quit smoking, 0 opposite).

  - Effect of an increase in the minimum wage on the probability of working (0 not working, 1 working).

- Recall that the (conditional) expectation of a binary variable (i.e., what we model in our regressions!) is the (conditional) probability that $y = 1$. To ensure that it's between 0 and 1 we might want to use nonlinear models that impose this.

- Other examples of non-linear models: count and censored data (see Cameron and Trivedi, Chapter 23).

- Also, we might like to allow for feedback from the past into future outcomes, relaxing strict exogeneity.

- For instance:

  - In a cigarette consumption model, we might need to include past consumption to capture addiction.

  - In a conflict regression, you might need to include past conflict in the equation. Conflict is very persistent and therefore being in conflict at $t-1$ increases a lot the probability of being in conflict at t.

  - Same happens with many economic variables, as they typically show large persistence.

This handout: advanced panel data models

- This handout focuses on estimation of

1) dynamic panel data models

2) nonlinear panel data models

- Topic is large!: this lecture presents a very short introduction

- We assume short panels (N large and T small) with time-invariant individual-specific effects, which may be:

  - Fixed.

  - Random (but remember that this not our preferred option).

Preview of the main results

- Dynamic panel data models

- All the estimators introduced in the previous section are inconsistent!

- Why? lags of the dependent variable are NOT strictly exogeneous.

- Within estimator: Nickell bias$\approx O(T^-1)$, (the problem disappears as $T \to \infty$ but we're assuming short panels!)

- New estimators: Difference GMM (also: System GMM)

Preview of the main results, II

Nonlinear panel data models

- If individual-specific effects are fixed and the panel is short:

  - Most models suffer from the incidental parameter problem.

  - Consistent estimation of slope parameters is possible for only a subset of nonlinear models (e.g., conditional logit, Poisson models. . .).

- If individual-specific effects are random:

  - Consistent estimation is possible for a wider range of models (e.g., Probit, Logit, Poisson, Tobit).

Dynamic Panel Data Models

Dynamic Panel Data Models

Autoregressive Models with Individual Effects

- Model Setup

Model:

$$y_{it} = c_i + X_{it}\beta + \rho\, y_{i,t-1} + v_{it}, \quad |\rho| < 1, \tag{22}$$

where:

- $i = 1, \ldots, N$ and $t = 1, \ldots, T$.
- Observed initial condition y_{i0}.
- c_i is an individual effect
- v_{it} satisfies error-components assumptions (serially uncorrelated, homokedastic). [Note: it's easy to relax the latter, always use robust options!]
- We can have additional lags of y_{it} in the regression: $y_{it-1}, y_{it-2}, \ldots$.

An example: Does democracy cause growth?

- Acemoglu et al. (JPE, 2019)
- Main question: does democracy cause growth?
- Main specification: dynamic linear panel data model.

DEMOCRACY DOES CAUSE GROWTH 57

$$y_{ct} = \beta D_{ct} + \sum_{j=1}^{p} \gamma_j y_{ct-j} + \alpha_c + \delta_t + \varepsilon_{ct}, \tag{1}$$

where y_{ct} is the log of GDP per capita in country c at time t and D_{ct} is our dichotomous measure of democracy in country c at time t. The α_c's denote a full set of country fixed effects, which will absorb the impact of any time-invariant country characteristics, and the δ_t's denote a full set of year fixed effects. The error term ε_{ct} includes all other time-varying unobservable shocks to GDP per capita. The specification includes p lags of log GDP per capita on the right-hand side to control for the dynamics of GDP, as discussed in Section I.

■ Why FE models?

“ [...], democracies differ from nondemocracies in unobserved characteristics, such as institutional, historical, and cultural aspects, that also have an impact on their GDP. As a result, cross-country regressions, as those in Barro (1996, 1999), could be biased and are unlikely to reveal the causal effect of democracy on growth.

■ Country FE are able to capture all time-invariant unobserved characteristics that countries have.

- Why dynamic models? “ [...] democratizations are, on average, preceded by a temporary dip in GDP. This figure depicts GDP dynamics in countries that democratized at year 0 relative to other countries that re- mained nondemocratic at the time. The pattern in this figure implies that failure to properly model GDP dynamics, or the propensity to de- mocratize based on past GDP, will lead to biased estimates of democracy on GDP.”

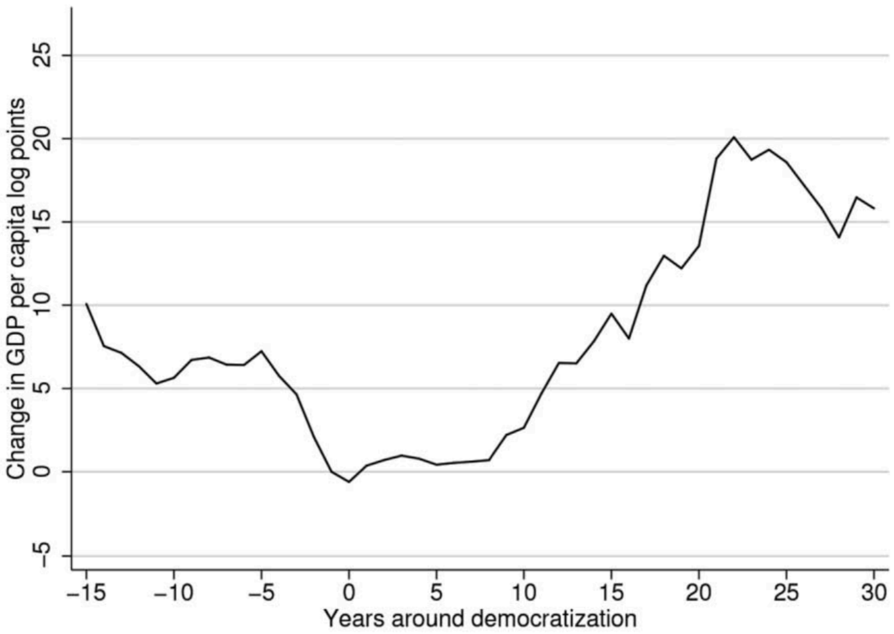

FIG. 1.—GDP per capita before and after a democratization. This figure plots GDP per capita in log points around a democratic transition relative to countries remaining nondemocratic in the same year. We normalize log GDP per capita to 0 in the year preceding the democratization. Time (in years) relative to the year of democratization runs on the horizontal axis.

An example: Does democracy cause growth?

- Their answer: yes!

> We provide evidence that democracy has a positive effect on GDP per capita. Our dynamic panel strategy controls for country fixed effects and the rich dynamics of GDP, which otherwise confound the effect of democracy. To reduce measurement error, we introduce a new indicator of democracy that consolidates previous measures. Our baseline results show that democratizations increase GDP per capita by about 20 percent in the long run. We find similar effects using a propensity score reweighting strategy as well as an instrumental-variables strategy using regional waves of democratization. The effects are similar across

Key difference between dynamic and static panels

- Main assumption in static panels: strict exogeneity.

- In dynamic panels, **strict exogeneity needs to be relaxed**

- Why?

- Notice that y_{it-1} is now a regressor, and that it contains v_{it-1}!

- Therefore, it's not strictly exogeneous: v_{it} is not uncorrelated with present, past and future values of y_{it-1}! $\rightarrow$ strict exogeneity fails!

- Key result:

- The estimators we've seen so far do not provide consistent estimators, even if c_i is random!

- Why?

- RE case: (i.e., c_i uncorrelated with X_{it})

  - y_{it-1} is correlated with c_i and hence with the composite error term $c_i + v_{it}$.

  - y_{it-1} is endogeneous (i.e., correlated with the error term)$\rightarrow$ OLS and the RE estimator are inconsistent!

FE case

- Set up: c_i correlated with X_{it}, within estimator is employed

  - y_{it} and v_{it} are correlated. Then y_{it-1} is correlated with v_{it-1} and y_{it-1} is correlated with $\bar{v}_i$.

  - within transformation: $\tilde{y}_{it-1} = y_{it-1} - \bar{y}_i$, $\tilde{v}_{it} = v_{it} - \bar{v}_i$

  - since $\tilde{v}_{it} = v_{it} - \bar{v}_i$ contains v_{it-1} and $\tilde{y}_{it-1}$ too, they are correlated!

Nickell bias:

- In short panels, the estimator of ρ is inconsistent and the bias is of the order (1/T), (Nickell, 1981)

- Meaning: The bias tends to disappear if T is large (i.e., the problem disappears in "long" panels).

- However, in short panels the bias can be large

  - In short panels, because the estimator of ρ is inconsistent, the estimator of β is so too.

  - The problem disappears in long panels, because the bias of the estimator of ρ goes to zero as

An estimator that does work:

Difference GMM estimation

- Consider the first difference transformation of the model:

First differences Model:

$$y_{it} - y_{i,t-1} = \rho(y_{i,t-1} - y_{i,t-2}) + (X_{it} - X_{i,t-1})'\beta + (v_{it} - v_{i,t-1})$$

This transformation:

- removes fixed effects (c_i).
- but endogeneity persists: $y_{i,t-1} - y_{i,t-2}$ is correlated with $v_{it} - v_{i,t-1}$! (why?).

Solution: Use instrumental variables (IV) to address endogeneity of $y_{i,t-1} - y_{i,t-2}$

Preliminary step: Anderson and Hsiao (1981) IV Estimator

- Problem: In general, finding good instruments can be a difficult task.

- **Key Idea of Anderson and Hsiao:** Use lagged values of y_{it} as instruments.

- More specifically: use $y_{i,t-2}$ as IV for the endogeneous variable: $y_{i,t-1} - y_{i,t-2}$

- Is $y_{i,t-2}$ a suitable IV?

  - Valid instrument: $y_{i,t-2}$ is uncorrelated with $(\epsilon_{it} - \epsilon_{i,t-1})$ if ϵ_{it} is serially uncorrelated.

  - "strong" instrument: $y_{i,t-2}$ is (typically) correlated with $(y_{i,t-1} - y_{i,t-2})$.

Efficiency

- AH is only fully efficient if $T = 2$.

- For $T > 2$, more instruments exist and considering them improves efficiency.

- Thus, estimators that consider more IVs are more efficient

- This leads to Difference GMM

Difference GMM: Arellano-Bond (1991) GMM Estimator

- Difference GMM: same idea:

  1) first differences model

  2) instrument the first differences of the dependent variable with further lags of that variable

  3) consider more IVs (all the available IVs!).

Difference GMM: Arellano-Bond (1991) GMM Estimator

Key Innovation: Use multiple lags of y_{it} as instruments.

$$z_{it} = [y_{i,t-2}, y_{i,t-3}, \ldots, y_{i1}, X'_{it}]$$

- Panel GMM exploits all available instruments for efficiency.
- Overidentified model (more IVs than endogeneous variables)
- Use two-stage least squares (2SLS) or GMM for estimation.

The Arellano–Bond estimator is given by:

$$\hat{\beta}_{AB} = \left(\sum_{i=1}^{N} X_i' Z_i W_N \sum_{i=1}^{N} Z_i' X_i \right)^{-1} \left(\sum_{i=1}^{N} X_i' Z_i W_N \sum_{i=1}^{N} Z_i' y_i \right), \qquad (1)$$

where:

- X_i is a $(T-2) \times (K+1)$ matrix with the tth row $(y_{i,t-1}, x_{it}')$, $t = 3, \ldots, T$,
- y_i is a $(T-2) \times 1$ vector with the tth row y_{it},
- Z_i is a $(T-2) \times r$ matrix of instruments:

$$Z_i = \begin{bmatrix} z'_{i3} & 0 & \cdots & 0 \\ 0 & z'_{i4} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & z'_{iT} \end{bmatrix},$$

with $z'_{it} = [y_{i,t-2}, y_{i,t-3}, \ldots, y_{i1}, x'_{it}]$. Lags of x_{it} or Δx_{it} can additionally be used as instruments.

- Two-stage least squares (2SLS) and two-step generalized method of moments (GMM) correspond to different weighting matrices W_N (see Section 22.2.3 in Cameron and Trivedi, for instance).

- GMM becomes 2SLS if $W_N = (Z'Z)^{-1}$

- GMM with "optimal" weights is the most efficient option.

Example, continued: democracy and growth

- Nickell bias is likely to be small (N=175 and T=50)

TABLE 2
EFFECT OF DEMOCRACY ON (LOG) GDP PER CAPITA

	WITHIN ESTIMATES				ARELLANO AND BOND ESTIMATES				
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(
Democracy	.973	.651	.787	.887	.959	.797	.875	.659	.
	(.294)	(.248)	(.226)	(.245)	(.477)	(.417)	(.374)	(.378)	(
Log GDP, first lag	.973	1.266	1.238	1.233	.946	1.216	1.204	1.204	.
	(.006)	(.038)	(.038)	(.039)	(.009)	(.041)	(.041)	(.038)	(
Log GDP, second lag		−.300	−.207	−.214		−.270	−.193	−.205	
		(.037)	(.046)	(.043)		(.038)	(.045)	(.042)	
Log GDP, third lag			−.026	−.021			−.028	−.020	
			(.028)	(.028)			(.028)	(.027)	
Log GDP, fourth lag			−.043	−.039			−.036	−.038	
			(.017)	(.034)			(.020)	(.033)	
p-value, lags 5–8				.565				.478	
Long-run effect of democracy	35.587	19.599	21.240	22.008	17.608	14.882	16.448	11.810	12.
	(13.998)	(8.595)	(7.215)	(7.740)	(10.609)	(9.152)	(8.436)	(7.829)	(8.

Other possibilities: System GMM

■ Problem: If y_{it} is highly persistent (ρ close to 1), then y_{it} "close" to being a random walk:

$$\Delta y_{it} \approx v_{it}$$

■ If this happens, Δy_{it-1} and y_{it-2} are "close" to being uncorrelated

■ This means that y_{it-2} is a weak instrument.

■ Blundell and Bond (1998): System GMM as solution to the above mentioned problem

- Combines differences and levels to increase efficiency.
- Additional moment conditions:

$$E[y_{i,s}(\epsilon_{i,t} - \epsilon_{i,t-1})] = 0, \quad s \leq t-2$$

Summary

- In dynamic panel data models:

- OLS, RE are inconsistent even if c_i is a RE!

- Fixed effects lead to Nickell bias, making FE inconsistent.

- IV methods (e.g., Anderson-Hsiao) address endogeneity but may lack efficiency.

- Arellano-Bond GMM leverages multiple instruments for more efficient estimation.

- Blundell-Bond System GMM improves efficiency under strong persistence (ρ close to 1) and small T.

Difference GMM in STATA

- Two functions: xtbond (native) and xtbond2 (Roodman, 2009)

- Nice paper to understand the details of the different options (Roodman, 2009):

Click here for Roodman's paper.

- In the acemoglu et al. case, (from their replication materials)

y: income; dem: dummy for democracy; yy*: year effects

xtabond2 y l.y dem yy* , gmmstyle(y, laglimits(2 .)) gmmstyle(dem, laglimits(1 .)) ivstyle(yy* , p) noleveleq robust nodiffsargan

- "gmm style" IVs: variables instrumented using lags (full explanation in Roodman's paper)

- "IV style": include here the exogeneous variables

Nonlinear Panel Data Models

Nonlinear Panel data models

- Recap: in the linear individual-specific effects model

$$y_{it} = c_i + X_{it}\beta + \varepsilon_{it}$$

- We model conditional means:

$$E(y_{it}|c_i, X_{it}) = c_i + X_{it}\beta$$

- Individual-specific parameters are additive and can be differenced out (within transformation, FD, etc..

- In nonlinear models, the conditional mean is a nonlinear function g

$$E(y_{it}|c_i, X_{it}) = g(c_i, X_{it})$$

- $g(.)$ can have different forms, we assume it to be known up to some parameters.
- For binary y_{it}, examples include:
  - Probit: $g(.) = \Phi(c_i + X_{it}\beta)$, where $\Phi(.)$ is the standard normal CDF.
  - Logit: $g(.) = \Lambda(c_i + X_{it}\beta)$, where $\Lambda(.)$ is the logistic CDF.

Incidental Parameters Problem

- Challenges:

- Individual-specific tems are not additive, cannot be differenced away.

- Estimation involves nuisance parameters $c_1, \ldots, c_N$: incidental parameters.

- The incidental parameters are inconsistently estimated as $N \to \infty$ if T is fixed (only T observations per parameter).

- This inconsistency contaminates the estimation of the common parameters, even though they use $NT \to \infty$ observations! It also results in inconsistent parameters.

A simple illustration

- Suppose $y_{it} \sim N[\alpha_i, \sigma^2]$.
- Maximum likelihood estimation yields:

$$\hat{\alpha}_i = \bar{y}_i, \quad i = 1, \ldots, N,$$

and

$$\hat{\sigma}^2 = \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} (y_{it} - \bar{y}_i)^2.$$

- It can be shown that as $N \to \infty$, $\hat{\sigma}^2 \approx \sigma^2 \frac{T-1}{T}$

■ Thus, $\hat{\sigma}^2$ is inconsistent for σ^2 as $N \to \infty$ in the short panel setting where T is fixed.

■ This inconsistency can be significant, with $\hat{\sigma}^2 \to_p 0.5\sigma^2$ when $T = 2$.

Incidental Parameter Problem in Panel data

- Consider inference when some parameters are common to all observations but there are an infinity of additional parameters that depend only on a finite number of observations.

- In the previous examples:

- β: common parameters.

- $c_1, \ldots, c_N$: incidental parameters, if T is fixed.

- The incidental parameters are inconsistently estimated as $N \to \infty$ (only T observations per parameter).

- This inconsistency contaminates the estimation of β, the common parameters, even though they use $NT \to \infty$ observations!

- In general if there is an incidental parameters problem, alternative estimation methods are needed that first eliminate the incidental parameters.

- For some popular models, most notably the panel probit model, there is no solution to the incidental parameters problem.

- No unified solution to the incidental parameters problem exists:

- Even where methods exist to consistently estimate β, these methods tend to be model specific

Solutions to the incidental parameter problem

The Conditional Likelihood

- For individual-specific effects panel models, if a sufficient statistic exists for the nuisance or incidental parameters (c_i), by conditioning on this sufficient statistic the nuisance parameter is eliminated.

- Recall: A statistic τ is called sufficient for a parameter θ if the distribution of the sample given τ does not depend on θ.

- By conditioning on this sufficient statistic the nuisance parameter c_i is eliminated. The resulting conditional density depends only on the common parameters, permitting consistent estimation.

- Key advantage:

- if the sufficient statistic for the incidental parameters exists, then it would be possible to:

1) compute the density of y conditional to the sufficient statistic,

2) this distribution won't depend on the incidental parameters (by definition of sufficient statistic)

3) One can use this distribution to estimate the other parameters by maximum likelihood (compute the likelihood function, maximize with respect to the key parameters etc)

- Key Challenge:

- Requires a sufficient statistic s_i, which exists only for a limited set of models (e.g., linear exponential family: normal, Poisson, binomial, gamma).

- Introducing regressors makes finding s_i even more challenging.

Binary outcome Data

- Now we consider a case for which the use of nonlinear models can be suitable: y_i is binary.

- Examples:
  - Employment status across several time periods.
  - Having a university degree or not
  - A country is in conflict or not

  ...

- Why use nonlinear models?

- In binary data: the conditional expectation of y_{it} is the conditional probability of y_{it}

- We can still use linear models: Linear Probability models

- But: we might obtain estimated values for y that are smaller than zero or larger than 1!

- The use of nonlinear models (i.e., Logit) ensures that estimated probabilities are in fact probabilities (i.e., between 0 and 1).

- The problem now is: how to estimate a nonlinear panel data model when y_it is binary.

- Solution: conditional Logit

Binary Models with Individual-Specific Effects

- Setup: y_{it} binary, FE suspected (i.e., unobserved effects, possibly correlated with the regressors.

- For simplicity: assume static models (extensions for dynamic models are possible)

- Fixed effects estimation:

  - Possible for the logit model.
  - Not possible for the probit model

- As mentioned before, a general solution to the incidental parameter problem doesn't exist.

Binary Models with Individual-Specific Effects

- Extend binary outcome model to panel data:

$$\Pr(y_{it} = 1 | X_{it}, \beta, c_i) = \Lambda(c_i + X_{it}\beta),$$

  where Λ is the Logistic function, $\Lambda(z) = \frac{e^z}{1+e^z}$

- (Assuming independence), joint density for $y_i = (y_{i1}, \ldots, y_{iT})$:

$$f(y_i | X_i, c_i, \beta) = \prod_{t=1}^{T} \Lambda(c_i + X_{it}\beta)^{y_{it}} (1 - \Lambda(c_i + X_{it}\beta))^{1-y_{it}}.$$

Fixed Effects Logit Model: Conditional Logit

- Joint density for y_i:

$$f(y_i|c_i, X_i, \beta) = \prod_{t=1}^{T} \frac{\exp(c_i + X_{it}\beta)^{y_{it}}}{1 + \exp(c_i + X_{it}\beta)}.$$

- Problem: incidental parameters: $c_1, \ldots, c_N$

- Can we find a sufficient statistic: YES!

- The statistic $\sum_{t=1}^{T} y_{it}$ is sufficient for c_i i.e., the distribution f(.) doesn't depend on c_i

- Then: Condition on $\sum_{t=1}^{T} y_{it} = d$ to eliminate c_i:

- It follows that:

$$f(y_i|c_i, X_{it}, \beta, \sum_{t=1}^{T} y_{it} = d) = f(y_i| \sum_{t=1}^{T} y_{it} = d, X_{it}, \beta),$$

- Therefore, one can maximize this likelihood function that doesn't have an incidental parameter problem to obtain estimates for β

STATA example

- Example from Allison's 2009 book *Fixed Effects Regression Models*.
- Data are from the National Longitudinal Study of Youth (NLSY).
- The data set has 1151 teenage girls who were interviewed annually for 5 years beginning in 1979.
- Problem: estimate the conditional probability of being in poverty (binary outcome)
- The variables are:

- `id` is the subject id number and is the same across each wave of the survey.
- `year` is the year the data were collected in. 1 = 1979, 2 = 1980, etc.
- `pov` is coded 1 if the subject was in poverty during that time period, 0 otherwise.
- `age` is the age at the first interview.
- `black` is coded 1 if the respondent is black, 0 otherwise.
- `mother` is coded 1 if the respondent currently has at least 1 child, 0 otherwise.
- `spouse` is coded 1 if the respondent is currently living with a spouse, 0 otherwise.
- `school` is coded 1 if the respondent is currently in school, 0 otherwise.
- `hours` is the hours worked during the week of the survey.

```
. use https://www3.nd.edu/~rwilliam/statafiles/teenpovxt, clear
. xtlogit pov i.mother i.spouse i.school hours i.year, fe nolog
note: multiple positive outcomes within groups encountered.
note: 324 groups (1,620 obs) dropped because of all positive or
      all negative outcomes.

Conditional fixed-effects logistic regression   Number of obs     =      4,135
Group variable: id                              Number of groups  =        827

                                                Obs per group:
                                                              min =          5
                                                              avg =        5.0
                                                              max =          5

                                                LR chi2(8)        =      97.28
Log likelihood  = -1520.1139                    Prob > chi2       =     0.0000

------------------------------------------------------------------------------
         pov |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
    1.mother |   .5824322   .1595831     3.65   0.000      .269655    .8952094
    1.spouse |  -.7477585   .1753466    -4.26   0.000    -1.091431   -.4040854
    1.school |   .2718653   .1127331     2.41   0.016     .0509125    .4928181
       hours |  -.0196461   .0031504    -6.24   0.000    -.0258208   -.0134714
             |
        year |
          2  |   .3317803   .1015628     3.27   0.001      .132721    .5308397
          3  |   .3349777   .1082496     3.09   0.002     .1228124     .547143
          4  |   .4327654   .1165144     3.71   0.000     .2044013    .6611295
          5  |   .4025012   .1275277     3.16   0.002     .1525514     .652451
------------------------------------------------------------------------------
```

- Interpretation:

- The note “multiple positive outcomes within groups encountered” is a warning that you may need to check your data, because with some analyses there should be no more than one positive outcome. In the present case, that is not a problem, i.e. there is no reason that respondents cannot be in poverty at multiple points in time.

- The note “324 groups (1620 obs) dropped because of all positive or all negative outcomes” means that 324 subjects were either in poverty during all 5 time periods or were not in poverty during all 5 time periods. Fixed-effects models are looking at the determinants of within-subject variability. If there is no variability within a subject, there is nothing to examine. Put another way, in the 827 groups that remained, sometime during the 5 year period the subject went from being in poverty to being out of poverty; or else switched from being out of poverty to being in poverty.

- If poverty status were something that hardly ever changed across time, or if very few people were ever in poverty, there would not be many cases left for a fixed effects analysis. Even as it is, more than a fourth of the sample has been dropped from the analysis.

■ Interpretation of the coefficients: it's useful to compute the "odds ratio"

```
. xtlogit, or

Conditional fixed-effects logistic regression   Number of obs     =      4,135
Group variable: id                              Number of groups  =        827

                                                Obs per group:
                                                              min =          5
                                                              avg =        5.0
                                                              max =          5

                                                LR chi2(8)        =      97.28
Log likelihood  = -1520.1139                    Prob > chi2       =     0.0000

------------------------------------------------------------------------------
         pov |         OR   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
    1.mother |   1.790388   .2857157     3.65   0.000     1.309513    2.447848
    1.spouse |   .4734266   .0830137    -4.26   0.000     .3357355    .6675871
    1.school |    1.31241   .1479521     2.41   0.016     1.052231    1.636923
       hours |   .9805456   .0030891    -6.24   0.000     .9745098    .9866189
             |
        year |
          2  |   1.393447   .1415223     3.27   0.001     1.141931    1.700359
          3  |   1.397909   .1513231     3.09   0.002     1.130672    1.728308
          4  |   1.541515   .1796087     3.71   0.000      1.22679    1.936979
          5  |   1.495561   .1907255     3.16   0.002     1.164802    1.920242
------------------------------------------------------------------------------
```

- Interpretation: The OR for mother is 1.79. This means that, if a girl switches from not having children to having children, her odds of being in poverty are multiplied by 1.79.

(these are teenagers at the start of the study, so having a baby while you are still very young is not good in terms of avoiding poverty.)

- Conversely, if a girl switches from being unmarried to married, her odds of being in poverty get multiplied by .47, i.e. getting married helps you to stay out of poverty.

- Being in school multiplies the odds of poverty by 31 percent, while each additional hour you work reduces the odds of poverty by 2 percent.

Conclusion

- Binary outcome panel models require careful modeling of individual effects

- Fixed effects logit models provide consistent estimation under conditional likelihood.