Longitudinal (Panel and Time Series Cross-Section) Data

Nathaniel Beck
Department of Politics
NYU
New York, NY 10012
nathaniel.beck@nyu.edu
http://www.nyu.edu/gsas/dept/politics/faculty/beck/beck_home.html

Jan. 2004

What is longitudinal data?

Observed over time as well as over space.

Pure cross-section data has many limitations (Kramer, 1983). Problem is that only have one historical context.

(Single) time series allows for multiple historical context, but for only one spatial location.

Longitudinal data - repeated observations on units observed over time

Subset of *hierarchical data* — observations that are correlated because there is some tie to same unit.

E.g. in educational studies, where we observe student i in school u. Presumably there is some tie between the observations in the same school.

In such data, observe $y_{j,u}$ where u indicates a unit and j indicates the j'th observation drawn from that unit. Thus no relationship between $y_{j,u}$ and $y_{j,u'}$ even though they have the same first subscript. In true longitudinal data, t represents comparable time.

Types of longitudinal data

- "Panel study" (NES, PSID, Congressional election outcomes by CD and year)
- Often use panel data as a single "enriched" cross-section, with info on prior behavior
- "Time-Series–Cross-Section" (political economy data on 15 OECD nations observed annually)
- Event history data
- Dyad year design in IR

- Data combining different surveys taken at different times (eg Markus article on Kramer)
- Rolling Cross-Section (Canadian Election Study)
- "Pseudo Panel" (group respondents by cohort) based on "Repeated Cross Section Data" (eg Family Expenditure Surveys)
- Binary dependent variable - estimation of transition matrix - Markov process

Texts

- Hsaio, *Analysis of Panel Data*, 2nd ed.
- Baltagi, *The Econometric Analysis of Panel Data*, 2nd ed.
- Matyas and Sevestera, *The Econometrics of Panel Data*, 1996 (handbook)

Panels vs TSCS data

Logically TSCS data looks like panel data, but panels have large number of cross-sections (big N) with each unit observed only a few times (small T); TSCS data has reasonable sized T and not very large N. For panel data, asymptotics are in N, T is fixed. For TSCS data, asymptotics in T, N is fixed.

This distinction is critical. Many of the panel methods are designed to deal with what is known as the "incidental parameters" problem, that is, as the number of parameters goes to ∞, one loses consistency. This is a problem only for panel, not TSCS data.

Furthermore, with small T there is no hope of saying anything about the time series structure of the data; with "biggish" T there is.

We also care about the units in TSCS data; they are states or countries. We do not usually care about the units in panel models; they are just a sample, and we care about the population parameters, not the sample. Thus "random effects" makes sense for panel models, whereas fixed effects make sense for TSCS data.

ALWAYS KEEP THIS DISTINCTION IN MIND

Statistical Issues - Estimation Technique

- OLS
- GLS (and FGLS)
- Full ML (and condition, or REML)
- How examine the likelihood
- Current - find mode and make asymptotic curvature assumptions
- Bayesian MCMC - explore entire likelihood - same church, different sect)

Issues that always arise in longitudinal data

- How model non-independent observations?
  - Repeated observations on same unit are seldom independent

- Assumption of independence should be tested, not assumed
  - So standard likelihood trick of breaking up likelihood of the sample into product of individual likelihoods does not work EASILY
- How model homogeneity?
  - Complete heterogeneity (area studies)
  - Complete homogeneity (econometrics)
  - Both positions silly, what is good compromise? Fixed effects?

Notation

The generic cross-national panel model we consider has the form:

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

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

Assessing Heterogeneity and simple fixes

Good idea to look at plots, at least of dv by unit (country)
Nice is box plot by unit
For Garrett data, with depvar being unem, stata command is

graph box unem, over(country)

Fixed Effects

Equation 1 assumes that all countries are fit by same model. And easy (though not trivial and also probably not enough) of a fix is to simply adjoin to the equation country specific intercepts α_i. These are simply dummy variables added to the OLS and so cause no estimation problems. Note that each effect is estimated with T observations, so no problem

as long as $T \to \infty$, even with large (very large) N. We return to this issue later.
Can test for whether need fixed effects by standard F-test, just compare the SSEs in usual way between Equation 1 and the specification with all the country dummies (that is, an F-test on the specifications with and without the dummy variables).
Some leave in only significant country dummies, but that is probably less than best practice (there are large se's on the dummies, so how much do you care if $t = 1.6$ or 1.7?
Note that all the fixed effects do is shift each country's regression line up and down, but leaves all regression lines parallel.
If fixed effects are needed in the model (that is, intercepts vary by unit), and you exclude, you will have specification error (omitted variable bias), very serious if a) the unit effects are non-trivial and b) the unit variables are correlated with the x's in the model.
For next while things are the same with and without fixed effects, so assume we have tested for them and if needed, they are just included as other iv's when they should be.

What it means to include FE's

It is not benign to include FE's. The standard results on partitioned regression show that the inclusion of FE's is equivalent to regressing the unit centered y's on the unit centered x's. Thus all the "between unit" variation in the variables is taken up by the FE's.
This is a particular problem for TSCS data, where the IV's may change very slowly. Of course if any IV does not change over time in any given unit, it is completely colinear with the dummy variable for that country.
But if we have, as we often do, variables that change very slowly (such as institutional measures), then FE's will essentially wipe them out.
Fortunately, in such cases, since the IV's already look like FE's, we will find on testing that we do not need the FE's. But with a huge amount of data (as in Green's "Dirty Pool"), we may find that standard tests reject the null that FE's do not belong. If that rejection is marginal, and if we have interesing IV's that do not vary a lot by time (but do vary by unit), then we might decide not to use FE's.
Remember, the harm of excluding FE's is a function of how much the tests indicate they should be included.
In the end, would we rather explain Germany by the name "Germany" or some structure like central bank independence?

Heterogeneity of more than intercepts

Classic Test
The classic test for pooling is to take Equation 1 as the null with the alternative being

complete heterogeneity, that is,

$$Ho : y_{i,t} = \mathbf{x}_{i,t}\beta + \epsilon_{i,t} \tag{2}$$

$$H1 : y_{i,t} = \mathbf{x}_{i,t}\beta_i + \epsilon i, t \tag{3}$$

that is, $H0 : \beta_i = \beta$.
The test of this is the standard F test, that is, take the difference of SSE's of the two models, correct for df, and put in the usual F ratio.
Note that if, say, $k = 5$ and $N = 20$, there are 100 df in the numerator, which is a lot.
(Note: the F test may suggest homogeneity of all coefs because most coefs are homogeneous and they few that aren't can't overcome the loss of degrees of freedom from those that are.)
If you have one variable (call is z of interest, with the others $(\mathbf{x})$ being more or less controls, you might consider only testing the heterogeneity of that parameter, and then if it is heterogeneous, estimating the model

$$Ho : y_{i,t} = \mathbf{x}_{i,t}\beta + z_{i,t}\gamma_i + \epsilon_{i,t} \tag{4}$$

Obviously this need not be limited to one z. (Lagged dependent variables and controls for the worldwide economy might be good candidates for homogeneity *in some models*.
(Note: You may reject the basic null because one or two coefficients vary enough and the F-test is not very conservative)
One also might worry that the F with so many degrees of freedom in the numerator will not pick a very parsimonious specification (the critical F value approaches one as the df in both numerator and denominator get large, that is, the F test becomes maximize $\bar{R}^2$.
While this would take us far afield, a criteria that picks more parsimonious models and that is liked by applied worked is the BIC or Schwarz criteria. (All criteria are f(SSE)+penalty for lack of parsimony, penalty is roughly $\frac{k}{N}$ for F-like criteria, much higher (roughly $\frac{k \ln N}{N}$) for Bayesian criteria.

Cross-Validation

One also might want to ask the data which countries do not fit the pooled specification. One way to do with would be via cross-validation.
In simplest form, c-v estimates a model leaving out one obs at a time, and then compares the prediction for that obs with the actual obs. This is the cross-sectional analogue of out-of-sample time series forecasting.
For large data sets, the cross-validated errors converge on the residuals (since dropping one obs hardly changes the estimated parms). For small data sets simple c-v is quite useful.
But one can also do "leave-out-k(%)" as well as "leave-out-one" c-v. This works nicely for TSCS data, where we can just leave one country out at a time, predict it, and then examine

the "forecast" errors. (CV is also used for model selection, and we could use this leave out one unit CV for this purpose also.)
This is done in the table for a model of Garrett's. Typical mean absolute forecast errors range from 1.2 to 2 (the unit is percent growth in GDP), except for Japan, which has a forecast error of 3.2% of GDP. Thus clearly Japan fits the basic specification much less well than any other OECD nation.

Table 1: Out of sample forecast errors (OLS) by country for Garrett model of economic growth in 14 OECD nations[a], 1966–1990

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

[a]No unit effects

Estimation - spherical errors

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

$$E(\epsilon_{i,t}\epsilon_{j,s}) = \begin{cases} \sigma^2 & \text{if } i = j \text{ and } s = t \\ 0 & \text{otherwise.} \end{cases} \tag{5}$$

But even if we like spherical errors, Equation 1 ASSUMES:

- all differences between units are accounted for by differences in the independent variables, that is, no "unmodeled heterogeneity"
- no effects of other units on each other - no spatial effects
- homogeneity (all units obey same equation) or "pooling"
- no dynamics (temporal dependence)

That is, G-M theorem assures us that under stated conditions that OLS of Equation 1 is optimal, but it doesn't tell us the Equation 1 has anything to do with how we think the world works.
Having said that, let us assume that Equation 1 represents our social science!

Non-spherical errors

It is unlikely that cross-national panel errors will meet the assumption of sphericality.
The usual OLS formula for standard errors will (always? often? sometimes?) provide misleading indications of the sampling variability of the coefficient estimates UNLESS THE ERRORS ARE SPHERICAL. The correct formula is given by the square roots of the diagonal terms of

$$\text{Cov}(\hat{\beta}) = (\mathbf{X}'\mathbf{X})^{-1}\{\mathbf{X}'\mathbf{\Omega}\mathbf{X}\}(\mathbf{X}'\mathbf{X})^{-1}. \tag{6}$$

OLS estimates this by

$$\widehat{\text{Cov}(\hat{\beta})}$$

$$= (\mathbf{X}'\mathbf{X})^{-1}\left\{\left(\frac{\sum_i \sum_t e_{i,t}^2}{NT-k}\right)\mathbf{X}'\mathbf{X}\right\}(\mathbf{X}'\mathbf{X})^{-1} \tag{7}$$

which then simplifies to the usual OLS estimate of the variance-covariance matrix of the estimates (the e's are OLS residuals). OLS standard errors are incorrect insofar as the middle terms in the two equations (in braces) differ.

ASIDE: What does it mean for standard errors to be incorrect, why not "inconsistent," and how would we know?

The OLS errors will be wrong if the errors show any of

- panel heteroskedasticity (Equation 16)
- contemporaneous correlation of the errors (Equation 17)
- serially correlated errors

Assuming observations are stacked by unit (not time), that is the first observation is unit 1, time 1, then unit 2, time 1, then unit 3, time 1, etc., we can write the VCV matrix for panel heteroskedastic and contemporaneously correlated (but temporally independent errors as)

$$\mathbf{\Omega} = \begin{pmatrix} \mathbf{\Sigma} & \mathbf{0} & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & \mathbf{\Sigma} & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{\Sigma} & \dots & \mathbf{0} \\ & & & \cdot & \\ & & & \cdot & \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \dots & \mathbf{\Sigma} \end{pmatrix} = \mathbf{\Sigma} \otimes \mathbf{I_N} \tag{8}$$

where

$$\mathbf{\Sigma} = \begin{pmatrix} \sigma_1^2 & \sigma_{1,2} & \sigma_{1,3} & \cdots & \sigma_{1,N} \\ \sigma_{1,2} & \sigma_2^2 & \sigma_{2,3} & \cdots & \sigma_{2,N} \\ & & & \cdot & \\ & & & \cdot & \\ \sigma_{1,N} & \sigma_{2,N} & \sigma_{3,N} & \cdots & \sigma_N^2 \end{pmatrix} \tag{9}$$

Models with temporally independent errors

NOTE: We deal with dynamics shortly.

For panel models with contemporaneously correlated and panel-heteroskedastic (but temporally independent) errors, $\mathbf{\Omega}$ is an $NT \times NT$ matrix block diagonal matrix with an $N \times N$ matrix of contemporaneous covariances, $\mathbf{\Sigma}$ (having typical element $E(\epsilon_{i,t}\epsilon_{j,t})$), along the diagonal. To estimate Equation 7 we need an estimate of $\mathbf{\Sigma}$. Since the OLS estimates of Equation 1 are consistent, we can use the OLS residuals from that estimation to estimate $\mathbf{\Sigma}$. Let $e_{i,t}$ be the OLS residual for unit i at time t. We can estimate a typical element of $\mathbf{\Sigma}$ by

$$\hat{\mathbf{\Sigma}}_{i,j} = \frac{\sum_{t=1}^{T} e_{i,t}e_{j,t}}{T}. \tag{10}$$

Letting $\mathbf{E}$ denote the $T \times N$ matrix of the OLS residuals, we can estimate $\mathbf{\Sigma}$ by

$$\hat{\mathbf{\Sigma}} = \frac{(\mathbf{E}'\mathbf{E})}{T} \tag{11}$$

and hence estimate $\mathbf{\Omega}$ by

$$\hat{\mathbf{\Omega}} = \frac{(\mathbf{E'E})}{T} \otimes I_T \tag{12}$$

where $\otimes$ is the Kronecker product.
We can then compute "Panel Correct Standard Errors" (PCSEs) by taking the square root of the diagonal elements of

$$(\mathbf{X'X})^{-1}\mathbf{X'}\left(\frac{\mathbf{E'E}}{T} \otimes \mathbf{I}_T\right)\mathbf{X}(\mathbf{X'X})^{-1}. \tag{13}$$

These still use the OLS estimates of $\hat{\beta}$ but provide correct reports of the variability of these estimates.

Generalized Least Squares

An alternative is GLS. If $\mathbf{\Omega}$ is known (up to a scale factor), GLS is fully efficient and yields consistent estimates of the standard errors. The GLS estimates of β are given by

$$(\mathbf{X'\Omega}^{-1}\mathbf{X})^{-1}\mathbf{X'\Omega}^{-1}\mathbf{Y} \tag{14}$$

with estimated covariance matrix

$$(\mathbf{X'\Omega}^{-1}\mathbf{X})^{-1}. \tag{15}$$

(Usually we simplify by finding some "trick" to just do a simple transform on the observations to make the resulting variance-covariance matrix of the errors satisfy the Gauss-Markov assumptions. Thus, the common Cochrane-Orcutt transformation to eliminate serial correlation of the errors is almost GLS, as is weighted regression to eliminate heteroskedasticity.)
The problem is that $\mathbf{\Omega}$ is never known in practice (even up to a scale factor). Thus an estimate of $\mathbf{\Omega}$, $\hat{\mathbf{\Omega}}$, is used in Equations 14 and 15. This procedure, FGLS, provides consistent estimates of β if $\hat{\mathbf{\Omega}}$ is estimated by residuals computed from consistent estimates of β; OLS provides such consistent estimates. We denote the FGLS estimates of β by $\tilde{\beta}$.
In finite samples FGLS underestimates sampling variability (for normal errors). The basic insight used by Freedman and Peters is that $\mathbf{X'\Omega}^{-1}\mathbf{X}$ is a (weakly) concave function of $\mathbf{\Omega}$. FGLS uses an estimate of $\mathbf{\Omega}$, $\hat{\mathbf{\Omega}}$, in place of the true $\mathbf{\Omega}$. As a consequence, the expectation of the FGLS variance, over possible realizations of $\hat{\mathbf{\Omega}}$, will be less than the variance, computed with the $\mathbf{\Omega}$. This holds even if $\hat{\mathbf{\Omega}}$ is a consistent estimator of $\mathbf{\Omega}$. The greater the variance of $\hat{\mathbf{\Omega}}$, the greater the downward bias.
This problem is not severe if there are only a small number of parameters in the variance-covariance matrix to be estimated (as in Cochrane-Orcutt) but is severe if there are a lot of parameters relative to the amount of data.

ASIDE: Maximum likelihood would get this right, since we would estimate all parameters and take those into account. But with a large number of parameters in the error process, we would just see that ML is impossible. That would have been good.

Panel Heteroskedasticity

If the errors follow the form

$$E(\epsilon_{i,t}\epsilon_{j,s}) = \begin{cases} \sigma_i^2 & \text{if } i=j \text{ and } s=t \\ 0 & \text{otherwise} \end{cases} \quad (16)$$

we have panel heteroskedasticity. It differs from simple heteroskedasticity in that error variances are constant within a unit.

The GLS correction for panel heteroskedasticity is to estimate σ_i^2 from the residuals in the obvious way and then use those estimates in a weighted least squares procedure.

The problem with this procedure is that it is basically weighting units by how well they fit the underlying regression, and so is simply downweighting those that fit poorly, clearly "improving" observed measures of fit.

Note how different this is from theoretically motivated weighted least squares. Panel weighted least squares is problematic, at best. Note that PCSEs correct the standard errors for panel heteroskedasticity.

Contemporaneously correlated errors

Errors are contemporaneously correlated if

$$E(\epsilon_{i,t}\epsilon_{j,s}) = \begin{cases} \sigma_{i,j}^2 & \text{if } s=t \\ 0 & \text{otherwise.} \end{cases} \quad (17)$$

Note here the huge number of $\sigma_{i,j}^2$ to be estimated in GLS. The GLS procedure here, often known as Parks, thus provides horrible standard errors, off by several hundred percent unless $T >> N$. (Note: Parks doesn't work, that is matrices not invertible, if $T < N$.

Parks should be avoided unless either N is very small or T is very large. (Note that Parks is the same as Zellner's Seemingly Unrelated Regressions. Fortunately SUR is usually applied to a small number of equations with many time points per equation.)

Spatial ideas

There may be some relationship between units, with a bigger relationship between "nearby"

units. Nearby could be geographical or, say, measured by trade.
Simplest thing would be, in a model of OECD political economy, to add overall OECD economic growth into a model of individual country GDP growth. Better would be to take the performance of related economies, using say the trade-weighted average of growth in all trading partners. This causes no econometric problems whatsoever (so long as the errors are spherical).
This is Garrett's OECD demand variable (though a more careful trade weighting would have been better).

Spatial Ideas to Improve Parks

The problem with Parks is that there are too many estimated covariances in the error matrix. We could parameterize those covariances, say by assuming that the contemporaneous correlation of the errors for units i and j is just $\lambda d(i, j)$ where $d(i, j)$ is the distance between the two units. This would enable us to use FGLS with only one additional parameter, and hence the estimates obtained should have good properties and the standard errors should be reasonably accurate (getting better as T increases).
Note that the distance measure need not be geographic, it could be the interrelationship of the two economies as measured by trade.
We could also use this idea to improve on PCSEs, simply estimating λ and then putting the estimated variance covariance matrix of the errors into Equation 6.
(Note: it is the repeated time observations and the assumption that all covariances are contemporary that makes this work.)
For lots of interesting econometrics of spatial concepts, see Anselim, "Spatial Econometrics." We will not purse these further.

TSCS Models with temporally dependent errors

Dynamics in TSCS data is very similar to dynamics in time series data.
The errors in cross-national panel models may show temporal dependence as well as spatial dependence.
GLS assumption - errors are serially correlated (usually AR1 with annual data, could easily generalize/test)

$$\epsilon_{i,t} = \rho \epsilon_{i,t-1} + \mu_{i,t} \tag{18}$$

where the μ are independently distributed across time.
(Aside: Old fashioned silliness - assumed that the temporal dependence of the errors could be modeled as a *unit-specific* AR1) process

$$\epsilon_{i,t} = \rho_i \epsilon_{i,t-1} + \mu_{i,t} \tag{19}$$

Simulations show that it is better to impose the additional assumption that the ρ are homogeneous across units, that is, $\rho_i = \rho$. (The time series are too short to get a good estimate of each ρ_i, and there are too many parameters to estimate. In addition, the assumption of a common ρ seems reasonable, especially when we assume common β.)
We can then correct for this serial correlation in the usual manner. First run OLS, compute the serial correlation of the residuals(that is, regress the residuals on the lagged residuals and take the coefficient on the lagged residual as $\hat{\rho}$.) Then transform by subtracting $\hat{\rho}$ of the prior observation from the current one, and run OLS on the transformed observations.

Testing

We can test for serially correlated errors (with or without a lagged dependent variable) via the TSCS analogue of the standard Lagrange multiplier test. Just run OLS, capture the residuals, and regress the residuals on all the independent variables (including the lagged dependent variable if present) and the lagged residual. If the coefficient on the lagged residual is significant (with the usual t-test), we can reject the null of independent errors.
We could test for, and model, second order and higher serial correlation, but most of our TSCS data is annual, and so short lags are often okay. But if you have monthly or quarterly data, worry about longer lags.

Lagged Dependent Variable

Just as with any time series, we could also model dynamics with a lagged dependent variable. The arguments for doing so with TSCS are identical to doing so for simple time series.

- they make the dynamics part of the model, not just a nuisance
- there is seldom any reason to prefer serially correlated errors to a lagged dependent variable
- the LDV model assumes that the effects of all variables, measured and unmeasured (the unmeasured variables are just the error term) have impacts that die out exponentially, whereas the AR1 error model assumes that the measured variables (which presumably are those we care about) have only immediate impact but the unmeasured variables have impacts which die out exponentially, which at first blush seems like an odd assumption
- lagged dependent variables usually simple to estimate and interpret (even if testing indicates SMALL remaining serial correlation)
- only problem is if resulting errors show serial correlation, almost never a problem in practice

You can then use the above LM test to test for remaining serial correlation, hoping you won't find any!

With an LDV in the model, all the error correlations across time period usually pretty much disappear, making life easy.

Taking testing more seriously

Rather than wishing away any remaining serial correlation, or hoping that LDV's are good, you can remember the following:
The LDV model is (using ν to denote iid errors)

$$y_{i,t} = \mathbf{x}_{i,t}\beta + \phi y_{i,t} + \nu_{i,t} \tag{20}$$

while the AR1 error model is

$$y_{i,t} = \mathbf{x}_{i,t}\beta + \nu_{i,t} + \rho\epsilon_{i,t-1} \tag{21}$$

since

$$\epsilon_{i,t} = \nu_{i,t} + \rho\epsilon_{i,t-1} \tag{22}$$

and hence

$$y_{i,t} = \mathbf{x}_{i,t}\beta + \rho y_{i,t-1} - \mathbf{x}_{i,t-1}\beta\rho + \nu_{i,t} \tag{23}$$

so both LDV and AR1 errors are special cases of ADL model

$$y_{i,t} = \mathbf{x}_{i,t}\beta + \rho y_{i,t-1} - \mathbf{x}_{i,t-1}\gamma + \nu_{i,t} \tag{24}$$

Thus, following Hendry's advice to test from general to specific, we might start with the ADL setup, and then test the null that either $\gamma = 0$ or $\gamma = -\beta\rho$. If we go with the ADL setup, and find, via an LM test, that the errors in that model are IID, we can just do OLS on the ADL model.
The claim for LDV model is that in practice for TSCS data we do not find ourselves frequently preferring ADL to LDV, though no reason not to test for this (rather than assume it).

“Dirty Pool”

Note that if one has a dynamic model but ignores this, it looks a lot like a model with fixed effects. Take the worst case,

$$y_{i,t} = y_{i,t-1} + \; bx_{i,t}\beta + \epsilon_{i,t} \tag{25}$$

but you ignore the LDV (with coef of 1).
You would then have (ignoring the error term)

$$y_{i,1} = y_{1,0} + \mathbf{x}_{i,1}\beta \tag{26}$$

$$y_{i,2} = y_{i,1} + \mathbf{x}i,2\beta \tag{27}$$

$$= y_{i,0} + (\mathbf{x}_{i,2} + \mathbf{x}i,1)\beta \tag{28}$$

$$y_{i,3} = y_{i,0} + (\mathbf{x}_{i,3} + \mathbf{x}_{i,2} + \mathbf{x}i,1)\beta \tag{29}$$

so that one can see that the omitted $y_{i,0}$ mimics a fixed effect.
This is why Green et al. in “Dirty Pool” found such a strong need for fixed effects; they based this on an incorrect static model. When one estimates a dynamic model (with LDV), the need for fixed effects is marginal (varies with test).

Hurwicz bias

In dynamic PANEL models (with LDV) and fixed effects, it is well known that the fixed effects will downwardly bias the coefficient of the LDV. This is called Hurwicz bias (for Leonid Hurwicz, who noted this effect for time series half a century ago). It is sometimes called Nickell bias, for Stephen Nickell who rediscovered and applied to panels 20 years ago. This bias is VERY serious for short T panels. (With T=2, is about 50%). It is also known that the bias is $O(1/T)$. Thus for TSCS, at $T \to \infty$, the bias disappears.
It would take Monte Carlo studies to know how serious the impact is when $20 < T < 50$, but it is not likely that the bias here is more than 5-10%. But with smallish T and LDV and FE's, there could be a somewhat bigger problem. At this point this is uncharted territory for TSCS.
But the well known problems for PANEL models with FE's and LDV's are clearly much less serious for TSCS data.