Dynamic Panel Data Models

Peter Lindner

June 23, 2010

Contents

Motivation

Many economic issues are dynamic by nature and use the panel data structure to understand adjustment.
Examples:

- Demand (i.e. present demand depends on past demand)
- Dynamic wage equation
- Employment models
- Investment of firms; etc.

One way error component model

$$y_{it} = \delta y_{i,t-1} + x_{it}'\beta + u_{it} \quad i = 1, \ldots, N \quad t = 1, \ldots, T$$

where

$$u_{it} = \mu_i + \nu_{it}$$

- μ_i usual individual effects, when necessary $iid(0, \sigma_\mu^2)$
- ν_i usual error term $iid(0, \sigma_\nu^2)$
- independent of each other and among themselves

OLS

- y_{it} is correlated with μ_i
- $\rightarrow y_{i,t-1}$ is also correlated with μ_i
- $\Longrightarrow$ OLS is biased an inconsistent even if ν_{it} are not serially correlated

Fixed effects

- Within transformation sweeps out μ_i
- $(y_{i,t-1} - \bar{y}_{i,-1})$, where $\bar{y}_{i,-1} = \sum_{t=2}^{T} \frac{y_{i,t-1}}{T-1}$ is correlated with $(\nu_{it} - \bar{\nu_{i.}})$
- $\rightarrow$ FE estimator is biased, BUT consistent for $T \rightarrow \infty$ (not for $N \rightarrow \infty$)

Random effects

- Quasi-demeaning transforms the data to $(y_{i,t-1} - \theta \bar{y}_{i,-1})$ and accordingly for the other terms
- $(y_{i,t-1} - \theta \bar{y}_{i,-1})$ is correlated with $(u_{it} - \theta \bar{u_{i.}})$ because $\bar{u_{i.}}$ contains $u_{i,t-1}$ which is correlated with $y_{i,t-1}$
- $\rightarrow$ RE GLS estimator is biased

How to deal with this problem

There are several ways in the literature, e.g. correcting for the bias, system GMM estimation techniques, etc.

Idea of one approach: IV-estimation

- Take first differences to get rid of the individual effects
- Use all the past information of y_{it} for instruments
- and the structure of the error term to get consistent estimates

Differencing I

- Model

$$y_{it} = \delta y_{i,t-1} + u_{it}$$

- First difference

$$y_{it} - y_{i,t-1} = \delta(y_{i,t-1} - y_{i,t-2}) + (\nu_{it} - \nu_{i,t-1})$$

- First period we have observation on this model is period $t = 3$, we have

$$y_{i3} - y_{i,2} = \delta(y_{i,2} - y_{i,1}) + (\nu_{i3} - \nu_{i,2})$$

$\rightarrow y_{i1}$ is not correlated with the error and a valid instrument

Differencing II

- One period forward we have

$$y_{i4} - y_{i3} = \delta(y_{i3} - y_{i2}) + (\nu_{i4} - \nu_{i3})$$

$\rightarrow$ y_{i1} and y_{i2} are not correlated with the error and a valid instrument

- One more instrument for each of the following periods
- Define a matrix the contains all instruments of individual i

$$W_i = \begin{bmatrix} [y_{i1}] & 0 & \cdots & & 0 \\ 0 & [y_{i1}, y_{i2}] & 0 & \cdots & 0 \\ \vdots & & \ddots & & 0 \\ 0 & & \cdots & & [y_{i1}, \dots, y_{i,T-2}] \end{bmatrix}$$

- All instruments in the model: $W = [W_1', W_2', \dots, W_N']'$

Error term

- Variance-covariance matrix of the error

$$E[\Delta v_i \Delta v_i'] = \sigma_\nu^2 (I_N \otimes G)$$

where

$$G = \begin{bmatrix} 2 & -1 & 0 & \cdots & & 0 \\ -1 & 2 & -1 & 0 & \cdots & 0 \\ \vdots & & \ddots & & \vdots & 0 \\ 0 & \cdots & 0 & -1 & 2 & -1 \\ 0 & \cdots & 0 & 0 & -1 & 2 \end{bmatrix}$$

- Since the instruments are orthogonal to the error we have the moment condition (used later for GMM)

$$E[W_i' \Delta v_i] = 0$$

Consistent estimates

- Pre-multiplying the model with the matrix of all instruments gives
$$W'\Delta y = W'(\Delta y_{-1})\delta + W'\Delta \nu$$
- Performing GLS on this model gives the consistent one step estimator (Arellano and Bond, 1991)

$$\hat{\delta}_1 = [(\Delta y_{-1})'W(W'(I_N \otimes G)W)^{-1}W'(\Delta y_{-1})]^{-1}$$
$$x[(\Delta y_{-1})'W(W'(I_N \otimes G)W)^{-1}W'(\Delta y)]$$

Optimal GMM estimates

- It can be shown that the the optimal GMM estimator (la Hansen) for this model is the same formula except replacing

$$(W'(I_N \otimes G)W)$$

by

$$V_N = \sum_{i=1}^{N} W_i'(\Delta v_i)(\Delta v_i)' W_i$$

where the Δv are obtain from the residuals form the above explained estimation

- Two step Arellano and Bond (1991) estimator is then

$$\hat{\delta}_1 = [(\Delta y_{-1})' W(\hat{V}_N)^{-1} W'(\Delta y_{-1})]^{-1}$$
$$x[(\Delta y_{-1})' W(\hat{V}_N)^{-1} W'(\Delta y)]$$

The data

Illustration with Arellano-Bonds dataset (can be freely downloaded from the web)

- firm level employment (Arellano-Bond 1991:Some tests of specification for panel data: Monte Carlo evidence and an application to employment equations, Review of Economic Studies)
- 140 UK firms
- annual data 1976-1984
- unbalanced

Idea behind the estimations

- Hiring and firing workers is costly
- Employment should adjust with delay to changes in factors such as capital stock, wages, and output demand and on the difference between equilibrium employment level and the last years actual level

$\longrightarrow$ Dynamic model where lags of the dependent variable are also regressors

Summary Statistics

- Employment is at the firm level and output is at the industry level as a proxy for demand
- Estimations are done with the logarithm of the variables

Table: **Summary Statistics**

Statistics	Employment	Wage	Capital	Output
Mean	7.89	23.92	2.51	103.80
Standard Deviation	15.94	5.65	6.25	9.94
Minimum	.10	8.02	.02	86.9
Maximum	108.56	45.23	47.11	128.37

Source: http://www.stata-press.com/data/r10/abdata.dta

First naive approach - OLS

Stata comand:
reg n nL1 nL2 w wL1 k kL1 kL2 ys ysL1 ysL2 yr*

Variable	Coefficient	(Std. Err.)
nL1	1.045**	(0.034)
nL2	-0.077*	(0.033)
w	-0.524**	(0.049)
wL1	0.477**	(0.049)
k	0.343**	(0.026)
kL1	-0.202**	(0.040)
kL2	-0.116**	(0.028)
ys	0.433**	(0.123)
ysL1	-0.768**	(0.166)
ysL2	0.312**	(0.111)
year dummies are not reported		

Problems with OLS

- Lagged dependent variable is endogenous to fixed effects in the error term
- Estimates are inconsistent
- One can show that there is a positive correlation between regressor and error term
- Thus it inflates the coefficient for lagged employment by attributing predictive power to it that belongs to the fixed effect
- One should expect the true estimate to be lower

Fixed effects: disregarding the dynamic structure

Stata comand:
xtreg n nL1 nL2 w wL1 k kL1 kL2 ys ysL1 ysL2 yr*, fe

Variable	Coefficient	(Std. Err.)
L1 employment	0.733**	(0.039)
L2 employment	-0.139**	(0.040)
Wage	-0.560**	(0.057)
L1 wage	0.315**	(0.061)
Capital	0.388**	(0.031)
L1 capital	-0.081*	(0.038)
L2 capital	-0.028	(0.033)
Output	0.469**	(0.123)
L1 Output	-0.629**	(0.158)
L2 Output	0.058	(0.135)
year dummies are not reported		

Problems with Fixed Effects Model

- Purging out the individual effects does not eliminate dynamic panel bias, it essentially makes every observation of transformed y* endogenous to the error
- One cannot use previous lags as instruments
- Estimates are now biased downwards
- Reasonable estimates should therefore lie between these FE-and OLS estimates; i.e. between 1.045 and 0.733

Arellano-Bond (difference GMM)

Stata comand:
xtabond2 n L.n L2.n w L1.w L(0.2).(k,ys) yr*, gmm(L.n) iv(L2.n w L.w L(0.2).(k,ys) yr*) nolevel robust

Variable	Coefficient	(Std. Err.)
L1 employment	0.686**	(0.145)
L2 employment	-0.085	(0.056)
Wage	-0.608**	(0.178)
L1 wage	0.393*	(0.168)
Capital	0.357**	(0.059)
L1 capital	-0.058	(0.073)
L2 capital	-0.020	(0.033)
Output	0.609**	(0.173)
L1 Output	-0.711**	(0.232)
L2 Output	0.106	(0.141)
year dummies are not reported		

Problems with Arellano-Bond I - Application

- Estimate of lagged dependent variable is NOT in "credible range" between OLS and fixed effect estimator
- Problem?
- Blundell and Bond 1998 "do not expect wages and capital to be strictly exogenous in our employment application"
- Therefore one can instrument them too

Arellano-Bond with more instruments

Stata comand:
xtabond2 n L.n L2.n w L1.w L(0.2).(k ys) yr*, gmm(L.(n w k)) iv(L(0.2).ys yr*) nolevel robust small

Variable	Coefficient	(Std. Err.)
L1 employment	0.818**	(0.086)
L2 employment	-0.112*	(0.050)
Wage	-0.682**	(0.143)
L1 wage	0.656**	(0.202)
Capital	0.353**	(0.122)
L2 capital	-0.154†	(0.086)
L2 capital	-0.030	(0.032)
Output	0.651**	(0.190)
L1 Output	-0.916**	(0.264)
L2 Output	0.279	(0.186)
year dummies are not reported		

Problems with Arellano-Bond II - General

- Now the estimate is between the previously showed FE and OLS estimate
- Blundell and Bond (1998) show that difference GMM performs bad when y is close to a random walk because untransformed lags are weak instruments for transformed variables.
- Instead of transforming the regressors it transforms the instruments to make them exogenous to the fixed effect
- Additional assumption that first differences of instruments are uncorrelated with fixed effects is necessary.

Arelano-Bond (system GMM) - one step

Stata comand:
xtabond2 n L.n L(0.1).(w k) yr*, gmmstyle(L.(n w k)) ivstyle(yr*, equation(level)) robust small

Variable	Coefficient	(Std. Err.)
L1 employment	1.030**	(0.057)
L2 employment	-0.089*	(0.434)
Wage	-0.641**	(0.123)
L wage	0.534**	(0.148)
Capital	0.428**	(0.062)
L capital	-0.374**	(0.066)
year dummies are not reported		
Intercept	0.552**	(0.197)

Conclusions

- How can we estimate a dynamic model with panel data
- It is relatively complicated in theory but easy with stata
- One has to carefully check the results from stata, because it always gives estimates.

Literature

Econometric Analysis of Panel Data

Baltagi, Badi H., John Wiley & Sons, Ltd (2005); 3rd edition; especially Chapter 8

How to Do xtabond2: An Introduction to "Difference" and "System" GMM in Stata

Roodman David, Center for Global Development, Working Paper no. 103, revised version 2001