Notes on Dynamic Panel Data

Sometimes we are interested in series that are correlated over time, for instance we might want to know the persistence of the amount of dividend per share distributed to shareholders. If we observe firms' dividend payout over a lot of periods, we can rely on a time series analysis. In contrast, if we have multiple firms and a short time series, we must rely on panel data estimators.

Consider the following model:

$$y_{i,t} = \eta_i + \rho y_{i,t-1} + v_{i,t}; \quad t = 1, ..., T \quad i = 1, ..., N \tag{1}$$

where

- $y_{i,t}$ is the dollar amount of dividends per share at year t.
- $y_{i,t-1}$ is the dollar amount of dividends per share at year $t-1$.
- η_i represent some feature of the firm that is invariant through our observed time periods, such as the sector.
- $v_{i,t}$ is the shock to the dividend policy at time t.

Our objective is to estimate ρ. For that, we need to make some general assumptions:

A1 The population model is linear in parameters.

A2 We have a random sample from the cross section.

A3 We have sample variation in the dependent variable in a large number of individuals (firms).

A4 Shocks are not affected by past levels of the dependent variable neither by the individual-specific level of y: $\mathbb{E}(v_{i,t}|y_{i,0}, ..., y_{i,t-1}, \eta_i) = 0 \; \forall t$.

A5 The variance of the shocks does not depend on the past, present or future value of$y$ or the individual heterogeneity: $Var(v_{i,t}|\eta, Y) = \sigma_v^2$

A6 Shocks are not correlated across time periods: $cov(v_{i,t}, v_{i,s}) = 0 \; \forall s \neq t$

Assumption A4 indicates that firms in growing sectors might distribute less dividends to their shareholders but the variation of their dividend policy is similar to the one pursue by firms in more mature sectors. Assumptions A5 and A6 can be relaxed, but they are useful to ease the exposition.

In the following sections we check if the static panel estimators are still valid under assumptions A1-A6.

Pooled OLS

The Pooled OLS estimator consider that firms are more or less similar to each other. In particular,

A7 There is no individual-specific heterogeneity: $\eta_i = \eta \; \forall i$

In this case the estimator is exactly the OLS estimator:

$$\hat{\rho} = \frac{\frac{1}{N}\sum_{n=1}^{N}\frac{1}{T}\sum_{t=1}^{T}(y_{i,t} - \overline{y})(y_{i,t-1} - \overline{y_{-}})}{\frac{1}{N}\sum_{n=1}^{N}\frac{1}{T}\sum_{t=1}^{T}(y_{i,t-1} - \overline{y_{-}})^2}$$

where $\overline{y} = \frac{1}{N}\sum_{n=1}^{N}\frac{1}{T}\sum_{t=1}^{T} y_{i,t}$ and $\overline{y_{-}} = \frac{1}{N}\sum_{n=1}^{N}\frac{1}{T}\sum_{t=1}^{T} y_{i,t-1}$. Substituting $y_{i,t}$ according to (1), we obtain:

$$\hat{\rho} = \rho + \frac{\frac{1}{N}\sum_{n=1}^{N}\frac{1}{T}\sum_{t=1}^{T}(v_{i,t} - \bar{v})(y_{i,t-1} - \bar{y})}{\frac{1}{N}\sum_{n=1}^{N}\frac{1}{T}\sum_{t=1}^{T}(y_{i,t-1} - \overline{y_{-}})^2}$$

Since we have a random sample in the cross-section we can apply the Law of Large Numbers, roughly, we can substitute $\frac{1}{N}\sum_{n=1}^{N}$ by $\mathbb{E}$. Therefore,

$$plim\,\hat{\rho} = \rho + \frac{\mathbb{E}\frac{1}{T}\sum_{t=1}^{T}(v_{i,t} - plim\,\bar{v})(y_{i,t-1} - plim\,\bar{y})}{\mathbb{E}\frac{1}{T}\sum_{t=1}^{T}(y_{i,t-1} - plim\,\overline{y_{-}})^2}.$$

Note that

$$plim\,\bar{v} = \mathbb{E}\frac{1}{T}\sum_{t=1}^{T} v_{i,t} = \frac{1}{T}\sum_{t=1}^{T}\mathbb{E}(v_{i,t}) = 0,$$

therefore,

$$plim\,\hat{\rho} = \rho + \frac{\frac{1}{T}\sum_{t=1}^{T}\mathbb{E}(v_{i,t}y_{i,t-1})}{\frac{1}{T}\sum_{t=1}^{T}\mathbb{E}(y_{i,t-1} - plim\,\overline{y_{-}})^2}$$

Using A4, we have that $\mathbb{E}(v_{i,t}y_{i,t-1}) = 0 \Rightarrow plim\,\hat{\rho} = \rho$. That is, $\hat{\rho}$ is consistent. In Stata we can implement it using `regress`:

```
. reg div l.div

      Source |       SS           df       MS      Number of obs   =     1,995
-------------+----------------------------------   F(1, 1993)      =    692.56
       Model |  322.320804         1  322.320804   Prob > F        =    0.0000
    Residual |   927.54598     1,993  .465401897   R-squared       =    0.2579
-------------+----------------------------------   Adj R-squared   =    0.2575
       Total |  1249.86678     1,994  .626813833   Root MSE        =     .6822

------------------------------------------------------------------------------
         div |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         div |
         L1. |   .5746759    .021837    26.32   0.000     .5318501    .6175016
             |
       _cons |   .1878168   .0171578    10.95   0.000     .1541676    .2214659
------------------------------------------------------------------------------
```

While pooled OLS is consistent it requires η_i to be the same for all individuals. In our example it requires all firms to have similar levels of dividend payout. To relax it, in the static panel data set-up, we transform the model to account for heterogeneity in η_i by subtracting either the average of each individual's observation (FE estimator), or the prior observation (FD estimator).

Fixed effect estimator (Within groups)

To compute the FE estimator subtract from every observation the individual sample mean, which leads to the following model:

$$y_{i,t} - \frac{1}{T}\sum_{s=1}^{T} y_{i,t} = \rho\left(y_{i,t-1} - \frac{1}{T}\sum_{s=1}^{T} y_{i,t-1}\right) + v_{i,t} - \frac{1}{T}\sum_{s=1}^{T} v_{i,t}; \quad t = 1, ..., T \quad i = 1, ..., N \quad (2)$$

If we define $\ddot{y}$ as the variable minus its time series average ($\ddot{y} = y_{i,t} - \frac{1}{T}\sum_{s=1}^{T} y_{i,t}$), we obtain:

$$\ddot{y}_{i,t} = \rho\ddot{y}_{i,t-1} + \ddot{v}_{i,t} \quad t = 1, ..., T \quad i = 1, ..., N \quad (3)$$

Note that we do not have heterogeneity, therefore A7 is satisfied. Nonetheless A4 is no longer satisfied:

$$cov(\ddot{v}_t, \ddot{y}_{t-1}) \neq 0$$

$$cov(\ddot{v}_t, \ddot{y}_{t-1}) = \underbrace{cov(y_{i,t-1}, v_{i,t})}_{\to 0} - \frac{1}{T}\sum_{s=1}^{T} cov(v_{i,s}, y_{i,t-1})$$
$$- \frac{1}{T}\sum_{s=1}^{T} cov(v_{i,t}, y_{i,s-1}) - \frac{1}{T^2}\sum_{s=1}^{T}\sum_{t=1}^{T} cov(v_{i,s}, y_{i,t-1})$$

We have four terms, the first one (in green) is the usual pooled OLS result which equals zero under assumption A4. The other three are mechanically created by the FE estimator and they are not zero (and they share the sign). Consider the first one (in orange):

$$\sum_{s=1}^{T} cov(v_{i,s}, y_{i,t-1}) = \sum_{s=1}^{t-2} cov(v_{i,s}, y_{i,t-1}) + cov(v_{i,t-1}, y_{i,t-1}) + \sum_{s=t}^{T} cov(v_{i,s}, y_{i,t-1})$$

$$= \boxed{\sum_{s=1}^{t-2} \rho^{t-1-s} \sigma_{\mathbf{v}}^{\mathbf{2}}} + \underbrace{\sum_{s=1}^{t-2} cov(v_{i,s}, v_{i,t-1})}_{\to 0} + \boxed{\sigma_{\mathbf{v}}^{\mathbf{2}}} + \underbrace{\sum_{s=t}^{T} cov(v_{i,s}, y_{i,t-1})}_{\to 0} \neq 0 \text{ ✗}$$

The terms inside the squares are never zero. The second squared term (in red) appears because when we subtract the average of $y_{i,t}$ on the left hand side of equation (2), we are including the covariate $y_{i,t-1}$ which now is part of both sides of the equation leading to a bias. In a similar way, the first squared term results from including $y_{i,t-1}$ through the autorregresive nature of the process. The time series average includes $y_{i,t-1}$ but it also includes $y_{i,t+s}$ $s \geq 0$ which indirectly contain $y_{i,t-1}$ as they equal $y_{i,t+s} = \rho^{s+1} y_{i,t-1} + \sum_{j=0}^{s} \rho^j v_{i,t-s}$.

In conclusion, the fixed effect estimator is not consistent if we consider a dynamic panel data model. Note, however, that the bias reduces as the time dimension (T) increases; furthermore, it completely disappears in the limit ($T \to \infty$). The following table present the bias for different values of T and ρ:

Table 1: Bias

T/ρ	0.05	0.5	0.95
2	-0.52	-0.75	-0.97
3	-0.35	-0.54	-0.73
10	-0.11	-0.16	-0.26
15	-0.07	-0.11	-0.17

We observe that if the true ρ equals 0.05 and we only observe each individual for two years, we will actually estimate that $\hat{\rho} = 0.05 - 0.52 = -0.47$!

First-difference estimator

An alternative to the FE estimator is the FD estimator which we obtain by subtracting to each observation its lag:

$$y_{i,t} - y_{i,t-1} = \rho(y_{i,t-1} - y_{i,t-2}) + v_{i,t} - v_{i,t-1}; \quad t = 2, ..., T \quad i = 1, ..., N \tag{4}$$

which eliminates the fixed effect. We can write the model as:

$$\Delta y_{i,t} = \rho \Delta y_{i,t-1} + \Delta v_{i,t}; \quad t = 2, ..., T \quad i = 1, ..., N$$

where $\Delta y_{i,t} = y_{i,t} - y_{i,t-1}$. Similar to the FE case, we can check if assumption A4 is valid:

$$\begin{aligned} cov(\Delta v_{i,t}, \Delta y_{i,t-1}) &= \cancelto{0}{cov(v_{i,t}, y_{i,t-1})} - \color{red}{cov(v_{i,t-1}, y_{i,t-1})} - \cancelto{0}{cov(v_{i,t}, y_{i,t-2})} + \cancelto{0}{cov(v_{i,t-1}, y_{i,t-2})} \\ &= -\color{red}{\sigma_v^2} \neq 0 \end{aligned}$$

The term in red is not zero, therefore the FD estimator is not consistent. The problem arises because when we subtract the lagged value from both sides, we are introducing endogeneity in the explanatory variable.

Solution: Instrumental variables

If the explanatory variable is endogenous, one possible solution is to use an instrument. In the FD case we can use an "internal" instrument, that is, we do not need to include any extra variable. For instance, if $\rho \neq 0$, $y_{i,t-2}$ is a valid instrument for $\Delta y_{i,t-1}$ since it satisfies:

- Relevance: $cov(y_{i,t-2}, \Delta y_{i,t-1}) = (\rho - 1)Var(y_{i,t-2}) \neq 0 \quad \checkmark$
- Exogeneity: $cov(y_{i,t-2}, \Delta v_{i,t}) = (y_{i,t-2}, v_{i,t}) + (y_{i,t-2}, v_{i,t-1}) = 0 \quad \checkmark$

Using $y_{i,t-2}$ as an instrument is know as the Anderson-Hsiao estimator. In Stata, it can be implemented using `ivreg` to the transformed model. Besides $y_{i,t-2}$, $y_{i,t-s}$ $\forall s > 2$ are also valid instruments (and so are their linear combinations). The command `xtabond` in Stata exploits all possible lagged ys as instruments. We can further improve the estimation by incorporating other functions of y as its lagged first differences ($\Delta y_{t-2}, \Delta y_{t-3}, ...$) which are also valid instruments for Δy_{t-1}. We can exploit instruments in levels and first differences using the command `xtdpdsys`:

```
. xtdpdsys div

System dynamic panel-data estimation            Number of obs      =      1,995
Group variable: idfirm                          Number of groups   =        285
Time variable: fyear
                                                Obs per group:
                                                              min =          7
                                                              avg =          7
                                                              max =          7

Number of instruments =      28                 Wald chi2(1)       =      13.98
                                                Prob > chi2        =     0.0002
One-step results
------------------------------------------------------------------------------
         div |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         div |
         L1. |    .086519   .0231435     3.74   0.000     .0411586    .1318794
             |
       _cons |   .3625649   .0152013    23.85   0.000      .332771    .3923589
------------------------------------------------------------------------------
```

We observe that the pooled OLS presents a huge positive bias which is consistent with the problem of spurious correlation in time series regressions. We can also compare these results with the bias estimates obtained by the FE and FD estimators:

```
. xtreg div l.div,fe

Fixed-effects (within) regression               Number of obs     =      1,995
Group variable: idfirm                          Number of groups  =        285

R-sq:                                           Obs per group:
     within  = 0.0001                                         min =          7
     between = 0.9372                                         avg =        7.0
     overall = 0.2579                                         max =          7

                                                F(1,1709)         =       0.24
corr(u_i, Xb)  = 0.7024                         Prob > F          =     0.6267
```

div	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
div						
L1.	.0135511	.0278599	0.49	0.627	-.0410919	.0681942
_cons	.3886856	.0167021	23.27	0.000	.3559269	.4214443
sigma_u	.55974113					
sigma_e	.59841047					
rho	.46664834	(fraction of variance due to u_i)				

```
. reg d.div d.(l.div),noconstant
```

Source	SS	df	MS			
Model	236.417246	1	236.417246	Number of obs	=	1,710
Residual	792.52132	1,709	.46373395	F(1, 1709)	=	509.81
				Prob > F	=	0.0000
Total	1028.93857	1,710	.60171846	R-squared	=	0.2298
				Adj R-squared	=	0.2293
				Root MSE	=	.68098

D.div	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
div						
LD.	-.5641255	.0249845	-22.58	0.000	-.6131289	-.5151221

As expected these coefficients are smaller than the one obtained through the instrumental variable approach. Furthermore, since T is fairly large ($T = 8$) the FE estimator is closer to the IV one than the FD estimator which is consistent with the bias derived previously.[1]

[1]The expected Nickell Bias in the FE case equals -0.16. The estimator of σ_v^2 equals 0.58, therefore we would expect a bias, in the FD case, of -0.58

Multivariate model

Similar to the time series case, if the dependent variable presents time dependence, it is important to include its lag, even if we are not interested in the nature of the dynamic process. For instance, with the same dataset we can estimate if a higher market-to-book ratio turns into a more generous dividend payout:

```
. xtdpdsys div MtB

System dynamic panel-data estimation          Number of obs      =      1,995
Group variable: idfirm                        Number of groups   =        285
Time variable: fyear
                                              Obs per group:
                                                            min =          7
                                                            avg =          7
                                                            max =          7

Number of instruments =      29               Wald chi2(2)       =      15.40
                                              Prob > chi2        =     0.0005
One-step results
------------------------------------------------------------------------------
         div |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         div |
         L1. |   .0853524   .0231612     3.69   0.000     .0399572    .1307476
             |
         MtB |   .0093779   .0080433     1.17   0.244    -.0063866    .0251425
       _cons |   .3367964   .0267823    12.58   0.000      .284304    .3892888
------------------------------------------------------------------------------
```

According to this estimation, there is no statistically significant relationship between market-to-book ratio and dividend payout. In contrast, if we do not account for the dividend payout dynamics, we estimate a significant coefficient whose magnitude almost doubles the previous one:

```
. xtreg div MtB,fe

Fixed-effects (within) regression               Number of obs     =      2,280
Group variable: idfirm                          Number of groups  =        285

R-sq:                                           Obs per group:
     within  = 0.0046                                         min =          8
     between = 0.0125                                         avg =        8.0
     overall = 0.0080                                         max =          8

                                                F(1,1994)         =       9.20
corr(u_i, Xb)  = 0.0293                         Prob > F          =     0.0025

------------------------------------------------------------------------------
         div |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         MtB |   .0160389   .0052877     3.03   0.002      .005669    .0264088
       _cons |   .3282767   .0189653    17.31   0.000     .2910827    .3654707
-------------+----------------------------------------------------------------
     sigma_u |  .52713506
     sigma_e |  .57772206
         rho |  .45430977   (fraction of variance due to u_i)
------------------------------------------------------------------------------
F test that all u_i=0: F(284, 1994) = 6.65                   Prob > F = 0.0000
```

Supplementary Reading: "Biases in Dynamic Models with Fixed Effects", Stephen Nickell Econometrica, Vol. 49, No. 6 (Nov.,1981), 1417-1426.