1 Estimation of Persistent Dynamic Panel Data. Motivation

Consider the following Dynamic Panel Data (DPD) model

$$y_{it} = y_{it-1}\rho + x_{it}\beta + \mu_i + v_{it} \tag{1.1}$$

with $i = \{1, 2, ..., N\}$ denoting the individual in a cross-section and $t = \{1, 2, ..., T\}$ denoting the time, which is assumed shorter than the cross-sectional dimension. By construction the lag-dependent variable y_{it-1} is correlated with the unobserved individual specific effect μ_i. The autoregressive parameter it assumed close or equal to one. The ordinary least square estimators in levels is inconsistent and upward bias (see Nickell (1981)) as the lagged dependent variable y_{it-1} is positively correlated with the error term $\mu_i + v_{it}$. The within groups estimator may also be inconsistent especially if the number of time periods is small compared with the number of individuals in the data as the transformed lagged dependent variable and the transformed error term are correlated. The reason is that in shorter panels the correlations of the leading observations which are negative dominate other correlations that are positive so that the correlation between the transformed lagged dependent variable and the transformed error term are negative, inducing downward bias. Alternative, likelihood based estimators can be used as they were developed, but they are highly dependent on the assumptions made on the distribution of the initial observation y_{i0} as one can make assumptions about: a) the stochastic nature of this observation, b) the correlation of this observation with the individual fixed effects, c) the stationarity of this initial observations. Therefore, using a likelihood estimator that misspecifies the process of the initial observation will induce also inconsistencies. An alternative to this methods is to use instrumental variables which require weaker assumptions on the initial conditions. In this context a transformation based on first differences is useful, firstly to eliminate the fixed effects from the equation (1) in order to obtain valid moment conditions and secondly by doing a first difference $(y_{it} - y_{it-1})$ one will not introduce all realizations of the past disturbances into the transformed error term, which will leave room on using some lagged variables for the construction of valid moment conditions. Additional moment conditions can be constructed depending on what it is assumed about the correlation between the x_{it} and the two components of the transformed error term $(v_{it} - v_{it-1})$. Therefore, the assumptions made about x_{it} are very important. The observables can be assumed to be correlated with: a) earlier shocks and not current shocks, b) current shocks but not past shocks, c) current shocks and past shocks but not future shocks, d) current shocks, past shocks and future shocks and e) nonstochastic (strictly exogenous) in other words not correlated with past, present or future shocks. Additional assumptions between the x_{it} and the fixed effects can be exploited to construct additional moment conditions and added to the moment conditions specified for the model in first difference:

$$\Delta y_{it} = \Delta y_{it-1}\rho + \Delta x_{it}\beta + \Delta v_{it},$$

where $\Delta y_{it} = y_{it} - y_{it-1}$, $\Delta y_{it-1} = y_{it-1} - y_{it-2}$, $\Delta x_{it} = x_{it} - x_{it-1}$ and $\Delta v_{it} = v_{it} - v_{it-1}$. When the data is highly persistent as it is in our case, one can exploit this feature and add additional moment conditions based on lagged differences Δy_{it-1} which can be used as valid instruments for the level equation (1). However, this feature can be used under the stationarity assumption of the initial observation y_{i0}, which requires

$$E(y_{i0} - \frac{\mu_i}{1-\rho})\mu_i = 0,$$

for all i, which implies that the initial observation do not deviate systematically from the stationary value $\frac{\mu_i}{1-\rho}$. Consequently, in the case of persistent data, the estimation does not depend only on first-differenced equations (2). Alternatively, one can also use moment conditions based on level equation (1). This is particularly important for very persistent data as the instruments that are available for the equation (2) are more likely to be weak (which may induce finite sample bias) and cannot be used for the identification of the ρ when we have unit root. Therefore, we need to have instruments that are not weak to reduce this potential bias.

For identification purposes, the above conditions require a large cross-sectional dimension. However, in our case, we can exploit only marginally the above mentioned properties as the cross-sectional dimension of the data is not large and may even become small when the data used is based on the information from only a limited number of countries.

To address the later problem, we are considering also estimators that are considering the finite aspects of the data.

1.1 Estimation

There is no estimator that can address all the problems of our data so, we consider in the analysis a set of estimators that are exploying different properties of the data such as: a) the finite sample properties: 1) the error correction based estimators and 2) the least square dummy variables estimators; b) the persistency of the data: 1) the system GMM estimators and, at the limit, 2) the cointegration based estimators.

- The error correction 2 SLS estimator (EC2SLS) proposed by Baltagi and Li (1992) is designed to improve the finite sample properties of the variance-covariance estimator resulting in finite sample efficiency gains that help to make better inference. However, the estimator is not helpful if the true model is dynamic in nature. The estimator is using a different projection matrix than the G2SLS estimator. Using the notation from Baltagi and Li (1992) on the statically defined model:

$$y_{it} = Z'_{it}\beta + u_{it},$$

where Z_{it} is defined as an $n \times g$ vector that includes both, the vector of observations on the explanatory variables X_{it} and a set of endogenous variables, $u_{it} = \mu_i + v_{it}$ with $\mu_i \approx iid(0, \sigma^2_{\mu)})$, $v_{it} \approx iid(0, \sigma^2_{v)})$, the consistent EC2SLS estimator takes the following form:

$$\widehat{\beta}_{EC2SLS} = (Z^{*\prime} P_A Z^*)^{-1} Z^{*\prime} P_A y^*,$$

where $P_A = A(A'A)^{-1}A'$, $A = [\widetilde{X}, \overline{X}]$ spans the set of instruments used by Balestra and Varadharajan-Krishnakumar (1987) ($X^* = \widetilde{X}/\sigma_v + \overline{X}/\sigma_1$), $\widetilde{X} = QX$ and $\overline{X} = PX$, $y^* = \Omega^{-1/2}y$, $Z^* = \Omega^{-1/2}Z$, with $\Omega^{-1/2} = \frac{P}{\sigma_1} + \frac{Q}{\sigma_v}$, $\sigma_1 = \sqrt{T\sigma^2_{mu} + \sigma^2_v}$. Further, $P = I_N \otimes \overline{J}_T$, with

$\overline{J}_T = J_T/T, Q = I_{NT} - P$, I_N is an identity matrix of dimension N, J_T is a matrix of ones of dimension T.

- The LSDV estimator of Kiviet (1995) for dynamic panels was evaluated in finite samples by Bun and Kiviet (2003). Bun and Kiviet (2003) showed that the first order term of the approximation evaluated at the true parameter values is can account for more than 90% of the actual bias. Using the model (1) presented above, the LSDV estimator of Kiviet (1995) for the true parameter $\theta = \{\rho, \beta\}$, is defined as:

$$\widehat{\theta}_{LSDV} = (W'M_sW)^{-1}W'M_sy,$$

where $M_s = S(I - D(D'SD)^{-1}D')S$ is symmetric and idempotent $NT \times NT$ matrix , $D = I_N \otimes i_T$ is a $NT \times N$ matrix of individual dummies, i_T is a $T \times 1$ vector of ones, $S = diag(S_i)$ a bloc-diagonal matrix of dimension $NT \times NT$ with S_i defining a $T \times T$ diagonal matrix that has as diagonal the $T \times 1$ vector $s_i = [s_{i1}...s_{iT}]$, that defines the dynamic selection rule (this selection rule uses only the observations that are usable for the dynamic model), in other words this selection rule is one if the pair (y_{it}, x_{it}) is observed and it is zero otherwise. The above estimator is biased, therefore an asymptotic biased correction composed of three parts is considered. However, it is proved in simulations that the asymptotic biased correction estimator performs well for a panel data with finite N dimension, data that is common to our application. The bias corrected LSDV estimator is therefore defined as:

$$\widehat{\theta}_{LSDVC} = \widehat{\theta}_{LSDVC} - \sum_{i=1}^{3} \widehat{Bias}_i \text{ with } i = 1, 2 \text{ and3.}$$

where the bias terms are defined in Bun and Kiviet (2003) for the balanced panels and in Bruno (2005)for the unbalanced panels $Bias_1 = c_1(\overline{T}^{-1})$, $Bias_2 = Bias_1 + c_2(N^{-1}\overline{T}^{-1})$ and $Bias_3 = Bias_2 + c_3(N^{-1}\overline{T}^{-2})$, where $\overline{T} = \sum_{i=1}^{N} T_i/N$ denotes the average group size. The simulation exercises of Bruno (2005) showed that for small cross-sections and not a large T the estimator works really well when the lagged dependent variable has a parameter around 0.8 or lower. In our dynamic panel data regression, the parameter of the lagged dependent regressor was estimated at about 0.8 or lower, which means that we should consider the LSDVC estimator as our preferred estimator. Additional robustness checks were done using GMM type estimators of Arrelano and Bond (2001), Blundell and Bond(1998) and cointegration based estimator of Pesaran and Smith (1995) and Pesaran, Shin and Smith (1999). It should be noted that the GMM type estimators are designed to work on micro panels, where the time-series T dimension is small compared with the cross-sectional N dimension. Within the class of the GMM estimators, the Arrelano and Bond (2001) estimator is not working well with persistent panels, while the Blundell and Bond(1998) estimator even if it is designed to work with persistent data, it requires a large cross-section. Alternatively, if we are willing to assume nonstationarity of our panel data, the cointegration based estimators of Pesaran et all (1999) are requiring a large time-series dimension.

- The GMM type estimator of Arrelano and Bond (2001) exploits the moment conditions:

$$E[Z_i' \Delta v_i] = 0 \text{ for } i = 1, 2, ..., N,$$

where $\Delta v_i = (\Delta v_{i3}, \Delta v_{i4}, ..., \Delta v_{iT})'$ and

$$Z_i = \begin{bmatrix} y_{i1} & 0 & 0 & ... & 0 & ... & 0 \\ 0 & y_{i1} & y_{i2} & ... & 0 & ... & 0 \\ . & . & . & ... & . & ... & . \\ 0 & 0 & 0 & ... & y_{i1} & ... & y_{i,T-2} \end{bmatrix}$$

where rows correspond to first-differenced equations of $t = 3, 4, ..., T$ for individual i. The asymptotically efficient GMM estimator based on this set of moment conditions minimizes the following objective function:

$$Q_N = (\frac{1}{N}\sum_{i=1}^{N}\Delta v_i Z_i)W_N(\frac{1}{N}\sum_{i=1}^{N}Z_i'\Delta v_i),$$

with the weighting matrix defined as

$$W_N = \left[\frac{1}{N}\sum_{i=1}^{N}(Z_i'\widehat{\Delta v}_i\widehat{\Delta v}_i'Z_i\right]^{-1},$$

and the $\widehat{\Delta v}_i$ are the consistent estimates of the first-differenced residuals obtained using a preliminary consistent method of estimation. This is known as a two step estimator. If the disturbances are homoscedastic by using the particular structure of the first-differenced model and asymptotic equivalent GMM estimator is obtained in one step using the alternative weighting matrix:

$$W_{N,a} = \left[\frac{1}{N}\sum_{i=1}^{N}(Z_i'HZ_i\right]^{-1},$$

where H is a (T-2) square matrix with 2's on the main diagonal, -1's on the first off-diagonal and zero elsewhere. This one-step estimator can be used as a consistent estimator for the two stage estimator. However, the two step estimator was shown in simulation to not produce efficiency gains even when the heteroscedasticity was significant, moreover, the simulation showed that the asymptotic standard errors of the two-step estimator are very small, implying big t-ratios.

2 References

Arellano, M. and S. Bond, 1991, Some Tests of Specification for Panel Data: Monte Carlo Evidence and an Application to Employment Equations. *Review of Economic Studies* 58: 277297.

Balestra, P. and J. Varadharajan-Krishnakumar, 1987, Full information estimations of a system of simultaneous equations with error components structure, *Econometric Theory* 3, 223.246.

Baltagi, B.H. and Q. Li, 1992, A note on the estimation of simultaneous equations with error components, *Econometric Theory* 8, 113.119.

Blundell, R. and S. Bond, 1998, Initial Conditions and Moment Restrictions in Dynamic Panel Data Models. J*ournal of Econometrics* 87: 115143.

Bruno, Giovanni S.F., 2005, Estimation and Inference in dynamic unbalanced panel data models with a small number of individuals, *The Stata Journal*, Volume 5 Number 4: pp. 473-500.

Bun, M. J. G. and J. F. Kiviet, 2003, On the diminishing returns of higher order terms in asymptotic expansions of bias. *Economics Letters* 79: 145152.

Kiviet, J. F. 1995, On Bias, Inconsistency and Efficiency of Various Estimators in Dynamic Panel Data Models. *Journal of Econometrics* 68: 5378.

Nickell, S. J., 1981, Biases in Dynamic Models with Fixed Effects. *Econometrica* 49: 14171426.

Pesaran, M. H. and Smith R., 1995, Estimating long-run relationships from dynamic heterogeneous panels, *Journal of Econometrics* Volume 68, Issue 1, July 1995, Pages 79-113.

Pesaran, M.H., Shin, Y., Smith, R.P., 1999, Pooled mean group estimation of dynamic heterogeneous panels. *Journal of the American Statistical Association* 94, 621 - 634.