Projection Estimators for Autoregressive Panel Data Models

Stephen Bond
Frank Windmeijer

THE INSTITUTE FOR FISCAL STUDIES
DEPARTMENT OF ECONOMICS, UCL
cemmap working paper CWP06/01

Projection Estimators for Autoregressive Panel Data Models

Stephen Bond[ab] and Frank Windmeijer[b*]
[a]Nuffield College, University of Oxford, Oxford, OX1 1NF
[b]CEMMAP, Institute for Fiscal Studies, London, WC1E 7AE

August, 2002

Abstract

In this paper we explore a new approach to estimation for autoregressive panel data models, based on projecting the unobserved individual effects on the vector of observations on the lagged dependent variable. This approach yields estimators which coincide with known generalised method of moments (GMM) estimators for models where stationarity is not imposed on the initial conditions and for models which satisfy mean stationarity. Our approach allows us to obtain a simple linear estimator for models which satisfy covariance stationarity, which although not fully efficient performs very well in simulations.
Key Words: Panel Data, Projections, Generalised Method of Moments
JEL CLassification: C13, C23

*Corresponding author. E-mail: f.windmeijer@ifs.org.uk

1. Introduction

In this paper we explore a new approach to estimation for autoregressive panel data models, based on projecting the unobserved individual effects on the vector of observations on the lagged dependent variable. This approach yields estimators which coincide with known generalised method of moments (GMM) estimators for models where stationarity is not imposed on the initial conditions (cf. Arellano and Bond, 1991) and for models which satisfy mean stationarity (cf. Blundell and Bond, 1998). Our approach allows us to obtain a simple linear estimator for models which satisfy covariance stationarity, which although not fully efficient performs very well in simulations.

Projection estimators for static panel data regression models with correlated individual effects were proposed by Chamberlain (1980, 1982, 1984). The basic idea is to consider the reduced form equations relating the endogenous variable in each period to the whole time series of observations on the exogenous variables. Parameters of interest are identified by imposing restrictions on this matrix of reduced form coefficients (the 'Π matrix'), and can be estimated using minimum distance methods. Recently Ruud (2000) has shown how the parameters of interest can be estimated more simply from a single least squares regression, in which the endogenous variable (y_{it}) is regressed on the exogenous variables (x_{it}) as well as *each* of the observations $x_{i1}, x_{i2}, ..., x_{iT}$, with identification achieved simply by imposing the constancy of parameters through time. This approach yields an estimator that can be implemented using a standard regression package.

Projection estimators for autoregressive panel data models with individual effects were suggested by Chamberlain (1980), and discussed in Sevestre and Trognon (1996) and Crépon and Mairesse (1996). These estimators are based on the projection of the individual effects on the initial conditions (y_{i1}). In this case the reduced form equations for each period relate y_{it} to the observation on y_{i1} only, and estimation again proceeds by imposing restrictions on this matrix of reduced form coefficients using minimum distance.

Our approach is based instead on the projection of the individual effects on the whole time series of observations on the lagged dependent variable. As in Ruud (2000), this allows the parameters of interest to be estimated simply from a single linear regression. In our case y_{it} is regressed on the lagged dependent variable ($y_{i,t-1}$) as well as each of the observations $y_{i1}, ..., y_{i,T-1}$. Consistent estimates of the autoregressive parameter of interest can be obtained from a simple two-stage least squares (2SLS) estimator, although since the model is over-identified (for $T > 3$) a more efficient linear GMM estimator is also available. These estimators can be implemented using standard regression packages for panel data such as DPD (Arellano and Bond, 1998), and TSP (Hall and Cummins, 1999).

We consider three different models for the initial conditions. In the first, no restrictions are imposed on the initial conditions except that they are uncorrelated with later shocks to the autoregressive process. In this case our projection estimator coincides with the first-differenced GMM estimator for autoregressive panel data models, developed in Holtz-Eakin, Newey and Rosen (1988) and Arellano and Bond (1991). In the second model, the initial conditions are assumed to satisfy mean stationarity. In this case our projection estimator coincides with the 'system' GMM estimator developed in Arellano and Bover (1995), Ahn and Schmidt (1995) and Blundell and Bond (1998). In the third model, the initial conditions are assumed to satisfy covariance stationarity. In this case we obtain a projection estimator that exploits the covariance stationarity restriction, and which is non-linear in the parameters of interest. We also suggest a simpler linear projection estimator for the covariance stationary model which, although not fully efficient, is shown to perform very well in simulations.

The remainder of the paper is structured as follows. Section 2 describes the three models we consider. Section 3 reviews the GMM estimators that have been proposed for these models. Section 4 introduces the projection estimators that we consider. Section 5 reports a Monte Carlo study which investigates the finite sample properties of these estimators, and presents some asymptotic variance comparisons. Section 6 concludes.

2. Models

Consider the simple dynamic AR(1) panel data model

$$\begin{aligned} y_{it} &= \alpha y_{i,t-1} + u_{it} \\ u_{it} &= \eta_i + v_{it}, \end{aligned}$$

for $i = 1, ..., N$ and $t = 2, ..., T$; N is large, T is fixed and $|\alpha| < 1$. The observations are independent across individuals and the error term satisfies

$$E(\eta_i) = 0, \ E(v_{it}) = 0 \quad \text{for } i = 1, ..., N \text{ and } t = 2, ..., T$$

and

$$E(v_{it} v_{is}) = 0 \quad \text{for } i = 1, ..., N \text{ and } t \neq s.$$

In this paper we consider three models that impose different restrictions on the initial conditions y_{i1} and/or the error term.

The first model, **Model 1**, only assumes that the v_{it} are uncorrelated with y_{i1}:

$$E(y_{i1} v_{it}) = 0 \quad \text{for } i = 1, ..., N \text{ and } t = 2, ..., T.$$

This model is similar to Model 1 of Alonso-Borrego and Arellano (1999).[1]

[1] Model 1 of Alonso-Borrego and Arellano (1999) makes the stronger assumption $E(v_{it} | y_{i1}, y_{i2}, ... y_{i,t-1}) = 0$.

The second model, **Model 2**, imposes an error components structure on the error term, and mean stationarity on the process, implying that

$$E(\eta_i v_{it}) = 0 \quad \text{for } i = 1, ..., N \text{ and } t = 2, ..., T$$

$$y_{i1} = \frac{\eta_i}{1-\alpha} + \varepsilon_i \quad \text{for } i = 1, ..., N$$

and

$$E(\varepsilon_i) = E(\eta_i \varepsilon_i) = 0 \quad \text{for } i = 1, ..., N.$$

See for example Arellano-Bover (1995) and Blundell-Bond (1998). Our Model 2 is the same as Model 2 in Alonso-Borrego and Arellano (1999).

Finally, **Model 3** further imposes homoskedasticity and covariance stationarity,[2] implying that

$$E\left(\eta_i^2\right) = \sigma_\eta^2 \quad \text{for } i = 1, ..., N \tag{2.1}$$

$$E\left(v_{it}^2\right) = \sigma_v^2 \quad \text{for } i = 1, ..., N \text{ and } t = 2, ..., T \tag{2.2}$$

$$E\left(\varepsilon_i^2\right) = \frac{\sigma_v^2}{1-\alpha^2} \quad \text{for } i = 1, ..., N. \tag{2.3}$$

3. GMM Estimation

3.1. Model 1

In Model 1, there are the following $(T-1)(T-2)/2$ linear moment conditions available for the estimation of α by GMM

$$E(y_{is}\Delta u_{it}) = 0 \quad \text{for } t = 3, \ldots, T \quad \text{and} \quad s = 1, \ldots, t-2, \tag{3.1}$$

[2] We impose homoskedasticity across individuals $i = 1, ..., N$ for simplicity, although only homoskedasticity of the v_{it} disturbances over time is strictly required for the results presented below.

where $\Delta u_{it} = u_{it} - u_{i,t-1} = \Delta y_{it} - \alpha \Delta y_{i,t-1}$, see for example Arellano-Bond (1991). Specifying the instrument set as

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

the GMM estimator minimises

$$\left(\frac{1}{N} \sum_{i=1}^{N} Z_i' \Delta u_i \right)' W_N \left(\frac{1}{N} \sum_{i=1}^{N} Z_i' \Delta u_i \right)$$

where $\Delta u_i = [\Delta u_{i3}, \Delta u_{i4}, \dots, \Delta u_{iT}]'$, and W_N is a positive definite weight matrix that converges to a matrix W as $N \to \infty$. Under homoskedasticity of the v_{it}, the optimal weight matrix which results in the minimum asymptotic variance of the GMM estimator for α, based on these moment conditions, is given by

$$W_N = \left(\frac{1}{N} \sum_{i=1}^{N} Z_i' H Z_i \right)^{-1}$$

where H is a $(T-2)$ square matrix which has 2's on the main diagonal, -1's on the first subdiagonals and zeros elsewhere. Under more general conditions, an optimal two-step estimator is based on the weight matrix

$$W_N = \left(\frac{1}{N} \sum_{i=1}^{N} Z_i' \widehat{\Delta u_i} \widehat{\Delta u_i}' Z_i \right)^{-1}$$

where $\widehat{\Delta u_i}$ are the residuals based on an initial consistent estimator for α.

Simulation studies which have investigated the finite sample properties of this estimator include Blundell-Bond (1998) and Alonso-Borrego and Arellano (1999), and this differenced GMM estimator has been used in many applied studies.

3.2. Model 2

In Model 2 there are the following extra $(T-2)$ linear moment conditions available

$$E\left(u_{it} \Delta y_{i,t-1}\right) = 0 \quad \text{for } t = 3, \dots, T, \tag{3.2}$$

see Arellano-Bover (1995), Ahn-Schmidt (1995) and Blundell-Bond (1998). The 'system' GMM estimator for α is obtained by stacking the residuals from the differenced and level equations, and extending the instrument matrix to

$$Z_i^S = \begin{bmatrix} Z_i & 0 & \cdots & 0 \\ 0 & \Delta y_{i2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \Delta y_{i,T-1} \end{bmatrix}.$$

There is no feasible optimal one-step weight matrix for this model, except for the special case in which the v_{it} are homoskedastic and $\eta_i = 0$ for $i = 1, ..., N$. In that case the optimal weight matrix is given by

$$W_N = \left(\frac{1}{N} \sum_{i=1}^{N} Z_i^{S\prime} A Z_i^S \right)^{-1}$$

where

$$A = \begin{bmatrix} H & C \\ C' & I_{T-2} \end{bmatrix}$$

with H defined above, I_{T-2} the identity matrix of order $(T-2)$, and C a $(T-2)$ square matrix with 1's on the main diagonal, -1's on the first lower subdiagonal, and zeros elsewhere.

The finite sample properties of this estimator are investigated in Blundell-Bond (1998) and Alonso-Borrego and Arellano (1999). Blundell-Bond (2000) provide an application to production function data, and Bond-Hoeffler-Temple (2001) provide an application to empirical growth models.

3.3. Model 3

In Model 3, there are $(T-2)$ additional linear moment conditions due to the homoskedasticity (through time) of v_{it}, given by

$$E\left(y_{it}u_{it} - y_{i,t-1}u_{i,t-1}\right) = 0 \ \text{ for } \ t = 3, ..., T. \tag{3.3}$$

Ahn and Schmidt (1997) derive the following non-linear moment condition which is valid under the further assumption (2.3) of covariance stationarity

$$E\left[y_{i1}^2 + \frac{y_{i2}\Delta u_{i3}}{1-\alpha^2} - \frac{u_{i3}u_{i2}}{(1-\alpha)^2}\right] = 0. \tag{3.4}$$

Recently, Kruiniger (2000) has shown that the non-linear moment condition (3.4) can be replaced by the linear moment condition

$$E\left[(1-\alpha)\left(\Delta y_{i2}\right)^2 + 2\Delta y_{i2}\Delta y_{i3}\right] = 0. \tag{3.5}$$

The full set of $0.5 \times T(T+1) - 2$ linear moment conditions for Model 3 consists then of (3.1), (3.2), (3.3) and (3.5). The GMM estimator for this model is obtained by again stacking the residuals from the differenced and level equations, augmented by the residual $\left(\Delta y_{i2}\right)^2 + 2\Delta y_{i2}\Delta y_{i3} - \alpha\left(\Delta y_{i2}\right)^2$. The instrument matrix is then given by

$$Z_i^{CS} = \begin{bmatrix} Z_i & 0 & \cdots & \cdots & 0 & 0 & 0 & \cdots & \cdots & 0 \\ 0 & 0 & \cdots & \cdots & 0 & -y_{i2} & 0 & \cdots & \cdots & 0 \\ 0 & \Delta y_{i2} & \cdots & \cdots & 0 & y_{i3} & -y_{i3} & \cdots & \cdots & \vdots \\ \vdots & \vdots & \ddots & & \vdots & 0 & y_{i4} & \ddots & & \vdots \\ \vdots & \vdots & & \ddots & 0 & \vdots & \vdots & \ddots & -y_{i,T-1} & \vdots \\ 0 & 0 & \cdots & 0 & \Delta y_{i,T-1} & 0 & 0 & \cdots & y_{iT} & 0 \\ 0 & 0 & \cdots & 0 & 0 & 0 & 0 & \cdots & 0 & 1 \end{bmatrix}.$$

Unlike the GMM estimators for Models 1 and 2, there is no feasible optimal one-step weight matrix for this model, even in the case when $\sigma_\eta^2 = 0$, and the v_{it} are homoskedastic.

We are not aware of any Monte Carlo study that has investigated the small sample properties of this GMM estimator for the covariance stationary model, or indeed of any applied study that has utilised the extra moment condition (3.5).

4. Projection Estimation

In this section we analyse the properties of a projection approach. The projection of the individual effects conditional on $y_{i1}, y_{i2}, ..., y_{i,T-1}$ is given by

$$Proj\left(\eta_i | y_{i1}, y_{i2,} ..., y_{i,T-1}\right) = \delta_1 y_{i1} + \delta_2 y_{i2} + ... + \delta_{T-1} y_{i,T-1}, \tag{4.1}$$

where $Proj\left(z|x\right)$ represents the population least squares projection of a variable z on a set of variables x, see Chamberlain (1982).

Projection estimators for static panel data regression models were proposed by Chamberlain (1980, 1982, 1984). One class of estimators that extend this approach to autoregressive models project η_i on the initial conditions y_{i1} only (see Chamberlain, 1980, Sevestre-Trognon, 1996, and Crépon-Mairesse, 1996). These estimators require non-linear restrictions to be imposed on the reduced form, using minimum distance techniques. Our approach allows estimates to be calculated by imposing simple restrictions in a 2SLS regression, and can be implemented easily using regression packages for panel data, such as DPD (Arellano and Bond, 1998) and TSP (Hall and Cummins, 1999). Our projection approach is similar in spirit to that suggested by Ruud (2000, pp. 630-635) for static panel data regression models.

4.1. Model 1

Model 1 does not impose any structure on the relationship between y_{i1} and η_i. Therefore there are no implicit restrictions on the δ parameters in (4.1). The projection estimator is obtained from specifying the model as

$$\begin{aligned} y_{it} &= \alpha y_{i,t-1} + \delta' \underline{y}_i^{T-1} + \widetilde{u}_{it} \quad \text{for } i = 1, ..., N \text{ and } t = 2, ..., T \\ \widetilde{u}_{it} &= \widetilde{\eta}_i + v_{it}, \end{aligned} \tag{4.2}$$

where $\delta = [\delta_1, \delta_2, \ldots, \delta_{T-1}]'$, $\underline{y}_i^{T-1} = [y_{i1}, y_{i2}, \ldots, y_{i,T-1}]'$ and $\widetilde{\eta}_i = \eta_i - \delta' \underline{y}_i^{T-1}$. The important point is that, unlike the original η_i, here $\widetilde{\eta}_i$ is uncorrelated with $\underline{y}_i^{T-1}$ by construction, so that suitably lagged levels of y_{it} can be used as instruments in (4.2) without further transformation.

We consider estimating (4.2) by 2SLS, using the sequential instruments Z_i^P defined as

$$Z_i^P = \begin{bmatrix} y_{i1} & 0 & 0 & \ldots & 0 & \ldots & 0 \\ 0 & y_{i1} & y_{i2} & \ldots & 0 & \ldots & 0 \\ 0 & 0 & 0 & \ddots & 0 & \ldots & 0 \\ 0 & 0 & 0 & \ldots & y_{i1} & \ldots & y_{i,T-1} \end{bmatrix}.$$

A necessary condition for the identification of α is that $T \geq 3$. As shown in the Appendix this 2SLS estimator of α is numerically identical to the differenced GMM estimator that uses the optimal weight matrix under homoskedasticity of v_{it}.[3] The model is over-identified when $T > 3$, and an efficient two-step estimator can be obtained using the weight matrix

$$W_N = \left(\frac{1}{N} \sum_{i=1}^{N} Z_i^{P\prime} \widehat{\widetilde{u}}_i \widehat{\widetilde{u}}_i' Z_i^P \right)^{-1},$$

where $\widehat{\widetilde{u}}_i$ are the one-step residuals. This two-step estimator of α is also numerically identical to the two-step differenced GMM estimator described in section 3.1, when the corresponding one-step estimator is used to estimate the weight matrix.

4.2. Model 2

Model 2 imposes $(T-2)$ restrictions on the δ parameters, which can easily be imposed as follows. Under the assumptions that specify Model 2, $E(y_{it}\eta_i)$ is

[3] The OLS estimator for α in equation (4.2) coincides with the familiar within groups estimator, as also shown in the Appendix.

constant over time, and here we define[4]

$$E\left(y_{it}\eta_i\right) = \kappa \text{ for } t = 1, ..., T.$$

Let ι be a vector of 1's of order $(T-1)$. The projection parameters δ are then given by

$$\begin{aligned}\delta &= E\left(\underline{y}_i^{T-1}\left(\underline{y}_i^{T-1}\right)'\right)^{-1} E\left(\underline{y}_i^{T-1}\eta_i\right) \\ &= \kappa E\left(\underline{y}_i^{T-1}\left(\underline{y}_i^{T-1}\right)'\right)^{-1} \iota.\end{aligned}$$

Letting $Y_{-1} = \left[\underline{y}_1^{T-1}, \underline{y}_2^{T-1}, \ldots, \underline{y}_N^{T-1}\right]'$, the model incorporating the restrictions is specified as

$$y_{it} = \alpha y_{i,t-1} + \kappa\iota'\left(\frac{1}{N}Y_{-1}'Y_{-1}\right)^{-1}\underline{y}_i^{T-1} + \widetilde{\eta}_i + v_{it}$$

and the parameters α and κ are estimated by 2SLS, using the same instruments Z_i^P. As shown in the Appendix, the resulting estimator of α is numerically identical to the system GMM estimator described in section 3.2, using the optimal initial weight matrix for the case when the v_{it} are homoskedastic and $\eta_i = 0$ for $i = 1, ..., N$. In order to calculate the correct (robust) asymptotic standard errors and optimal two-step GMM estimator, the residuals $\widehat{u}_{it} = y_{it} - \widehat{\alpha}y_{i,t-1}$ have to be used. In that case, the two-step projection estimator and two-step system GMM estimator are also numerically identical.[5]

[4] If the η_i were homoskedastic across individuals, we would have $\kappa = \sigma_\eta^2/(1-\alpha)$, where $\sigma_\eta^2 = E\left(\eta_i^2\right)$ as for Model 3.

[5] The addition of other regressors to the model, with a corresponding extension of the projection, does not alter these equivalence results. This is discussed in the Appendix.

4.3. Model 3

Under the assumptions that specify Model 3, $E\left(\underline{y}_i^{T-1}\left(\underline{y}_i^{T-1}\right)'\right)$ is given by

$$E\left(\underline{y}_i^{T-1}\left(\underline{y}_i^{T-1}\right)'\right)=\frac{\sigma_\eta^2}{(1-\alpha)^2}J_{T-1}+\frac{\sigma_v^2}{1-\alpha^2}\begin{bmatrix}1 & \alpha & \cdots & \alpha^{T-2}\\ \alpha & 1 & \ddots & \vdots\\ \vdots & \ddots & \ddots & \alpha\\ \alpha^{T-2} & \cdots & \alpha & 1\end{bmatrix}, \tag{4.3}$$

where J_{T-1} is a $(T-1)$ square matrix of ones. The projection parameters δ are here given by

$$\delta=\frac{\sigma_\eta^2}{1-\alpha}E\left(\underline{y}_i^{T-1}\left(\underline{y}_i^{T-1}\right)'\right)^{-1}\underline{s}.$$

Using (4.3) it follows that the elements of δ are related in the following way[6]

$$\delta=\begin{bmatrix}\tau\\ (1-\alpha)\tau\\ (1-\alpha)\tau\\ \vdots\\ (1-\alpha)\tau\\ \tau\end{bmatrix}, \tag{4.4}$$

where

$$\tau=\frac{\sigma_\eta^2}{1-\alpha}\left(\frac{(1-\alpha)^2}{\sigma_\eta^2\left((T-1)-\alpha\left(T-3\right)\right)+\sigma_v^2\left(1-\alpha\right)}\right).$$

Therefore, the specification that incorporates all the restrictions on the projection parameters δ arising from covariance stationarity is

$$y_{it}=\alpha y_{i,t-1}+\tau\left(y_{i1}+y_{i,T-1}\right)+\left(1-\alpha\right)\tau\sum_{s=2}^{T-2}y_{is}+\widetilde{\eta}_i+v_{it}. \tag{4.5}$$

The parameters α and τ can be estimated from (4.5) using a non-linear GMM estimator based on the moment conditions

$$E\left[Z_i^{P\prime}\left(\widetilde{\eta}_i+v_{it}\right)\right]=0 \quad \text{for } t=2,...,T.$$

[6] This suggests that a simple test of whether the series satisfy covariance stationarity can be obtained by testing these restrictions on the projection parameters δ in model (4.2). See Bond, Bowsher and Windmeijer (2001) and Bond and Windmeijer (2002) for a discussion of various test procedures for linear restrictions in dynamic panel data models.

The resulting estimators of α do not coincide with either of the GMM estimators for Model 3, described in section 3.3. A comparison of the asymptotic variances presented in the next section confirms that this projection estimator is not asymptotically efficient. The finite sample properties of these alternative estimators for Model 3 will be investigated in our Monte Carlo simulations.

This projection approach also allows consistent estimation of the parameter of interest α using a simpler linear 2SLS or GMM estimator in the model

$$y_{it} = \alpha y_{i,t-1} + \tau \left(y_{i1} + y_{i,T-1}\right) + \gamma \sum_{s=2}^{T-2} y_{is} + \widetilde{\eta}_i + v_{it}, \tag{4.6}$$

again using the instrument matrix Z_i^P. Clearly, there will be some further loss of efficiency from not exploiting the non-linear restriction relating γ to α and τ, the extent of which will also be investigated in the next section.

5. Monte Carlo and Asymptotic Variance Comparisons

In this section we investigate the performance of the projection estimators in the three models. As the finite sample properties of the estimators we obtain for Model 1 and Model 2 are relatively well known,[7] the main contribution of this study is to assess the relative performance of the projection estimators we have proposed for Model 3 and further to compare them with the linear GMM estimator for this model discussed in section 3.3.

The covariance stationary process is generated as follows:

$$\begin{aligned} y_{it} &= \alpha y_{i,t-1} + \eta_i + v_{it} \\ \eta_i &\sim N\left(0, \sigma_\eta^2\right), \quad v_{it} \sim N\left(0, \sigma_v^2\right) \\ y_{i1} &= \frac{\eta_i}{1-\alpha} + \varepsilon_i, \quad \varepsilon_i \sim N\left(0, \frac{\sigma_v^2}{1-\alpha^2}\right). \end{aligned}$$

[7] See, for example, Blundell and Bond (1998) and Alonso-Borrego and Arellano (1999).

Tables 1 and 2 report estimation results for $\sigma_\eta^2 = \sigma_v^2 = 1$, $T = 4$, various values of α, and $N = 100$ and $N = 500$ respectively. Tables 3 and 4 report results for $T = 7$. We report the means and standard deviations of the empirical distributions of the estimators in our experiments, as well as the means of the calculated asymptotic standard errors for these estimators, and their root mean squared errors (rmse). Reported summary statistics are based on 5000 replications. The one-step estimators are computed using the weight matrices described above, and hence can be obtained as linear 2SLS estimators in the relevant projection specifications, except for the non-linear projection estimator we consider for Model 3.

Results for Model 1 are very similar to those reported in Blundell-Bond (1998) and Alonso-Borrego and Arellano (1999), as the design and estimators are identical. Compared to these studies, results for Model 2 are slightly different, as the projection based 2SLS estimator is equivalent to the one-step system GMM estimator using the optimal weight matrix when v_{it} is homoskedastic and $\eta_i = 0$ for $i = 1, ..., N$, whereas the one-step system GMM estimator considered in the previous two studies used $\left(\frac{1}{N}\sum_i Z_i^{S\prime} Z_i^S\right)^{-1}$ as the weight matrix.

The reduction in bias, especially at high values of α, and the increase in precision of the estimators of α that impose the restrictions implied by the mean stationarity of the process, are apparent.[8] For both models, the estimated robust asymptotic standard errors are close to the observed standard deviations for the one-step estimators. For the two-step estimators, the estimated standard errors can be much smaller than the empirical standard deviations, especially when N is small and T is large.[9]

[8] The estimates for $\kappa = \frac{\sigma_\eta^2}{1-\alpha}$ in Model 2 are downward biased and estimated imprecisely, especially for high values of α and small sample size N.

[9] Windmeijer (2000) develops a finite sample correction for the variance of two-step GMM estimators that addresses this problem.

Table 1. $T = 4$, $N = 100$, $\sigma_\eta^2 = 1$, $\sigma_v^2 = 1$, mean, standard deviation, [mean of estimated s.e.], and *rmse,* 5000 replications

	Model 1		Model 2	
α	One-Step	Two-Step	One-Step	Two-Step
0.0	-0.0074	-0.0074	0.0171	0.0138
	0.1230	0.1274	0.1084	0.1008
	[0.1200]	[0.1169]	[0.1058]	[0.0858]
	0.1232	*0.1276*	*0.1097*	*0.1017*
0.3	0.2800	0.2829	0.3153	0.3189
	0.1814	0.1886	0.1289	0.1211
	[0.1793]	[0.1747]	[0.1320]	[0.1049]
	0.1824	*0.1894*	*0.1298*	*0.1225*
0.5	0.4547	0.4599	0.5151	0.5207
	0.2646	0.2753	0.1429	0.1344
	[0.2573]	[0.2503]	[0.1465]	[0.1151]
	0.2684	*0.2781*	*0.1436*	*0.1360*
0.8	0.4741	0.4545	0.8439	0.8346
	0.6836	0.6951	0.1520	0.1589
	[0.6990]	[0.6640]	[0.1688]	[0.1186]
	0.7572	*0.7762*	*0.1582*	*0.1626*
0.9	0.2505	0.2256	0.9612	0.9519
	0.8477	0.8777	0.1157	0.1540
	[1.0012]	[0.9210]	[0.1522]	[0.0917]
	1.0679	*1.1068*	*0.1308*	*0.1625*

For the covariance stationary Model 3, Tables 1 to 4 present estimation results for the projection estimators described in section 4.3. The estimator labelled 'Non-Linear' is the projection estimator where the full set of restrictions on δ arising from the structure of $E\left(\underline{y}_i^{T-1}\left(\underline{y}_i^{T-1}\right)'\right)$ are imposed, as in model (4.5). The estimator labelled 'Linear' is the simple linear projection estimator of model (4.6). For this model we also present results for the linear GMM estimator that exploits the additional moment conditions (3.3) due to homoskedasticity and (3.5) due to covariance stationarity of the initial observations.

Table 1 cnt'd. $T = 4$, $N = 100$, $\sigma_\eta^2 = 1$, $\sigma_v^2 = 1$, mean, standard deviation, [mean of estimated s.e.], and *rmse*, 5000 replications

	Model 3					
	Projection				GMM	
	Nonlinear		Linear		Linear	
α	One-Step	Two-Step	One-Step	Two-Step	One-Step	Two-Step
0.0	-0.0043	-0.0036	-0.0045	-0.0030	0.0059	0.0085
	0.0933	0.0963	0.0926	0.0954	0.1383	0.1018
	[0.0901]	[0.0848]	[0.0895]	[0.0861]	[0.1343]	[0.0777]
	0.0934	*0.0964*	*0.0927*	*0.0955*	*0.1384*	*0.1021*
0.3	0.2887	0.2921	0.2890	0.2920	0.3092	0.3151
	0.1047	0.1074	0.1048	0.1069	0.1482	0.1182
	[0.1033]	[0.0973]	[0.1032]	[0.0994]	[0.1477]	[0.0878]
	0.1053	*0.1076*	*0.1054*	*0.1072*	*0.1485*	*0.1192*
0.5	0.4853	0.4910	0.4864	0.4912	0.5099	0.5227
	0.1131	0.1169	0.1145	0.1180	0.1574	0.1322
	[0.1090]	[0.1031]	[0.1100]	[0.1062]	[0.1533]	[0.0892]
	0.1141	*0.1173*	*0.1153*	*0.1183*	*0.1577*	*0.1341*
0.8	0.7799	0.7894	0.7830	0.7900	0.8251	0.8372
	0.1181	0.1233	0.1224	0.1264	0.1474	0.1376
	[0.1143]	[0.1091]	[0.1179]	[0.1145]	[0.1430]	[0.0639]
	0.1198	*0.1237*	*0.1235*	*0.1268*	*0.1495*	*0.1425*
0.9	0.8798	0.8907	0.8818	0.8900	0.9352	0.9375
	0.1318	0.1350	0.1234	0.1272	0.1160	0.1125
	[0.1149]	[0.1099]	[0.1197]	[0.1164]	[0.1089]	[0.0409]
	0.1333	*0.1353*	*0.1247*	*0.1276*	*0.1212*	*0.1186*

Table 2. $T = 4$, $N = 500$, $\sigma_\eta^2 = 1$, $\sigma_v^2 = 1$, mean, standard deviation, [mean of estimated s.e.], and *rmse,* 5000 replications

	Model 1		Model 2	
α	One-Step	Two-Step	One-Step	Two-Step
0.0	-0.0013	-0.0013	0.0036	0.0028
	0.0550	0.0554	0.0505	0.0429
	[0.0546]	[0.0543]	[0.0497]	[0.0408]
	0.0550	*0.0554*	*0.0506*	*0.0429*
0.3	0.2974	0.2980	0.3024	0.3047
	0.0816	0.0822	0.0639	0.0543
	[0.0808]	[0.0804]	[0.0633]	[0.0518]
	0.0816	*0.0822*	*0.0639*	*0.0545*
0.5	0.4923	0.4939	0.5014	0.5058
	0.1140	0.1153	0.0720	0.0625
	[0.1145]	[0.1139]	[0.0716]	[0.0601]
	0.1142	*0.1154*	*0.0720*	*0.0628*
0.8	0.7438	0.7442	0.8033	0.8029
	0.3030	0.3071	0.0819	0.0733
	[0.3000]	[0.2979]	[0.0855]	[0.0687]
	0.3082	*0.3122*	*0.0820*	*0.0733*
0.9	0.6387	0.6194	0.9204	0.9145
	0.6375	0.6625	0.0899	0.0959
	[0.6380]	[0.6202]	[0.0995]	[0.0664]
	0.6889	*0.7194*	*0.0921*	*0.0969*

Table 2 cnt'd. $T = 4$, $N = 500$, $\sigma_\eta^2 = 1$, $\sigma_v^2 = 1$, mean, standard deviation, [mean of estimated s.e.], and *rmse*, 5000 replications

	Model 3					
	Projection				GMM	
	Nonlinear		Linear		Linear	
α	One-Step	Two-Step	One-Step	Two-Step	One-Step	Two-Step
0.0	-0.0008	-0.0006	-0.0009	-0.0005	0.0028	0.0019
	0.0418	0.0416	0.0415	0.0415	0.0628	0.0411
	[0.0411]	[0.0401]	[0.0408]	[0.0403]	[0.0619]	[0.0383]
	0.0418	*0.0416*	*0.0415*	*0.0415*	*0.0629*	*0.0411*
0.3	0.2981	0.2990	0.2982	0.2991	0.3036	0.3037
	0.0474	0.0474	0.0473	0.0473	0.0692	0.0473
	[0.0471]	[0.0460]	[0.0469]	[0.0463]	[0.0689]	[0.0441]
	0.0475	*0.0474*	*0.0473*	*0.0473*	*0.0693*	*0.0475*
0.5	0.4976	0.4991	0.4978	0.4990	0.5021	0.5061
	0.0511	0.0509	0.0514	0.0513	0.0733	0.0521
	[0.0500]	[0.0490]	[0.0503]	[0.0498]	[0.0730]	[0.0469]
	0.0512	*0.0509*	*0.0514*	*0.0513*	*0.0733*	*0.0525*
0.8	0.7980	0.8008	0.7986	0.8006	0.8072	0.8169
	0.0523	0.0528	0.0537	0.0541	0.0763	0.0625
	[0.0520]	[0.0514]	[0.0537]	[0.0533]	[0.0750]	[0.0444]
	0.0523	*0.0528*	*0.0537*	*0.0541*	*0.0767*	*0.0647*
0.9	0.8984	0.9013	0.8987	0.9007	0.9117	0.9199
	0.0761	0.0770	0.0550	0.0553	0.0688	0.0615
	[0.0521]	[0.0516]	[0.0544]	[0.0541]	[0.0688]	[0.0327]
	0.0761	*0.0770*	*0.0550*	*0.0553*	*0.0698*	*0.0646*

Table 3. $T = 7$, $N = 100$, $\sigma_\eta^2 = 1$, $\sigma_v^2 = 1$, mean, standard deviation, [mean of estimated s.e.], and *rmse*, 5000 replications

	Model 1		Model 2	
α	One-Step	Two-Step	One-Step	Two-Step
0.0	-0.0135	-0.0138	0.0371	0.0145
	0.0624	0.0689	0.0642	0.0585
	[0.0617]	[0.0540]	[0.0613]	[0.0402]
	0.0638	*0.0702*	*0.0742*	*0.0603*
0.3	0.2733	0.2742	0.3495	0.3239
	0.0788	0.0870	0.0744	0.0662
	[0.0775]	[0.0678]	[0.0725]	[0.0439]
	0.0832	*0.0908*	*0.0893*	*0.0704*
0.5	0.4539	0.4551	0.5616	0.5365
	0.0959	0.1076	0.0781	0.0746
	[0.0955]	[0.0835]	[0.0768]	[0.0459]
	0.1064	*0.1166*	*0.0995*	*0.0830*
0.8	0.6128	0.5916	0.8854	0.8656
	0.1784	0.2159	0.0581	0.0715
	[0.1721]	[0.1489]	[0.0575]	[0.0378]
	0.2586	*0.3000*	*0.1033*	*0.0971*
0.9	0.5190	0.4398	0.9725	0.9645
	0.2481	0.3241	0.0352	0.0516
	[0.2372]	[0.1967]	[0.0340]	[0.0235]
	0.4546	*0.5628*	*0.0806*	*0.0826*

Table 3 cnt'd. $T = 7$, $N = 100$, $\sigma_\eta^2 = 1$, $\sigma_v^2 = 1$,mean, standard deviation [mean of estimated s.e.], and *rmse*, 5000 replications

	Model 3					
	Projection				GMM	
	Non Linear		Linear		Linear	
α	One-Step	Two-Step	One-Step	Two-Step	One-Step	Two-Step
0.0	-0.0074	-0.0070	-0.0074	-0.0060	-0.0038	-0.0024
	0.0509	0.0563	0.0507	0.0563	0.0902	0.0653
	[0.0500]	[0.0411]	[0.0498]	[0.0415]	[0.0873]	[0.0373]
	0.0514	*0.0568*	*0.0513*	*0.0566*	*0.0903*	*0.0653*
0.3	0.2891	0.2918	0.2890	0.2927	0.2964	0.3004
	0.0543	0.0590	0.0544	0.0594	0.0986	0.0712
	[0.0536]	[0.0440]	[0.0536]	[0.0450]	[0.0962]	[0.0393]
	0.0554	*0.0596*	*0.0555*	*0.0599*	*0.0987*	*0.0712*
0.5	0.4875	0.4921	0.4873	0.4930	0.5022	0.5063
	0.0548	0.0601	0.0555	0.0610	0.1026	0.0762
	[0.0547]	[0.0449]	[0.0552]	[0.0464]	[0.1007]	[0.0389]
	0.0562	*0.0606*	*0.0570*	*0.0614*	*0.1026*	*0.0764*
0.8	0.7812	0.7870	0.7814	0.7885	0.8245	0.8275
	0.0540	0.0597	0.0568	0.0628	0.0909	0.0830
	[0.0545]	[0.0454]	[0.0570]	[0.0486]	[0.0893]	[0.0277]
	0.0572	*0.0612*	*0.0598*	*0.0638*	*0.0941*	*0.0874*
0.9	0.8791	0.8868	0.8790	0.8861	0.9386	0.9387
	0.0537	0.0728	0.0574	0.0631	0.0634	0.0632
	[0.0536]	[0.0450]	[0.0574]	[0.0492]	[0.0601]	[0.0147]
	0.0576	*0.0740*	*0.0611*	*0.0646*	*0.0742*	*0.0741*

Table 4. $T = 7$, $N = 500$, $\sigma_\eta^2 = 1$, $\sigma_v^2 = 1$,mean, standard deviation, [mean of estimated s.e.], and *rmse*, 5000 replications

	Model 1		Model 2	
α	One-Step	Two-Step	One-Step	Two-Step
0.0	-0.0024	-0.0024	0.0080	0.0016
	0.0285	0.0294	0.0287	0.0236
	[0.0283]	[0.0275]	[0.0281]	[0.0213]
	0.0286	*0.0295*	*0.0298*	*0.0236*
0.3	0.2944	0.2947	0.3107	0.3018
	0.0364	0.0374	0.0360	0.0266
	[0.0358]	[0.0349]	[0.0352]	[0.0237]
	0.0368	*0.0378*	*0.0375*	*0.0267*
0.5	0.4901	0.4907	0.5139	0.5038
	0.0445	0.0461	0.0395	0.0291
	[0.0446]	[0.0434]	[0.0399]	[0.0260]
	0.0456	*0.0470*	*0.0419*	*0.0293*
0.8	0.7539	0.7536	0.8249	0.8128
	0.0880	0.0926	0.0398	0.0378
	[0.0875]	[0.0852]	[0.0417]	[0.0295]
	0.0993	*0.1036*	*0.0470*	*0.0399*
0.9	0.7566	0.7418	0.9371	0.9233
	0.1518	0.1741	0.0311	0.0373
	[0.1491]	[0.1443]	[0.0316]	[0.0230]
	0.2088	*0.2352*	*0.0484*	*0.0439*

Table 4 cnt'd. $T = 7$, $N = 500$, $\sigma_\eta^2 = 1$, $\sigma_v^2 = 1$,mean, standard deviation, [mean of estimated s.e.], and *rmse*, 5000 replications

	Model 3					
	Projection				GMM	
	Non linear		Linear		Linear	
α	One-Step	Two-Step	One-Step	Two-Step	One-Step	Two-Step
0.0	-0.0008	-0.0007	-0.0008	-0.0004	-0.0006	0.0004
	0.0229	0.0237	0.0228	0.0236	0.0417	0.0240
	[0.0227]	[0.0216]	[0.0226]	[0.0217]	[0.0407]	[0.0209]
	0.0229	*0.0237*	*0.0228*	*0.0236*	*0.0417*	*0.0240*
0.3	0.2978	0.2986	0.2978	0.2987	0.3001	0.3005
	0.0246	0.0250	0.0246	0.0253	0.0462	0.0255
	[0.0244]	[0.0231]	[0.0244]	[0.0234]	[0.0457]	[0.0223]
	0.0247	*0.0251*	*0.0247*	*0.0253*	*0.0462*	*0.0255*
0.5	0.4977	0.4988	0.4977	0.4991	0.5009	0.5013
	0.0251	0.0255	0.0253	0.0259	0.0485	0.0263
	[0.0249]	[0.0236]	[0.0252]	[0.0242]	[0.0492]	[0.0227]
	0.0252	*0.0255*	*0.0254*	*0.0259*	*0.0485*	*0.0263*
0.8	0.7962	0.7981	0.7963	0.7984	0.8060	0.8069
	0.0246	0.0251	0.0260	0.0267	0.0518	0.0319
	[0.0247]	[0.0237]	[0.0259]	[0.0250]	[0.0518]	[0.0215]
	0.0249	*0.0252*	*0.0262*	*0.0268*	*0.0521*	*0.0326*
0.9	0.8956	0.8977	0.8953	0.8975	0.9168	0.9163
	0.0238	0.0287	0.0257	0.0265	0.0456	0.0380
	[0.0241]	[0.0233]	[0.0260]	[0.0252]	[0.0452]	[0.0159]
	0.0242	*0.0288*	*0.0261*	*0.0266*	*0.0486*	*0.0414*

The two projection estimators are found to be very similar in their performance. Both provide a significant improvement in root mean squared error compared to the estimators for Model 2, particularly at high values of α. For both estimators the one-step version is generally found to have a smaller variance than the two-step version in these experiments, particularly at $N = 100$. Again we find that the estimated standard errors are reliable for the one-step estimators.

The GMM estimator for Model 3 also provides an improvement over that for Model 2. In contrast to the two projection estimators for Model 3, there appears to be a serious loss in precision from not using the optimal weight matrix. However, even the optimal two-step GMM estimator is generally found to have a slightly larger small sample variance and root mean squared error than the projection estimators in these experiments. This is perhaps surprising given that this GMM estimator is asymptotically efficient in the class of estimators that make use of second moment information.[10]

To investigate this further, Figure 1 presents plots of the asymptotic standard deviations of these two-step estimators for Model 3, calculated for the design as in the Monte Carlo experiments with $T = 4$ and for various values of α. As expected, the two-step GMM estimator for Model 3 does have a slightly smaller asymptotic standard deviation than the non-linear projection estimator, with the difference diminishing for increasing values of α. The simple linear projection estimator has a slightly larger asymptotic standard deviation than the non-linear projection estimator, with the difference getting smaller for low values of α. Nonetheless, the results in Tables 1-4 suggest that this simple linear estimator performs as least as well as the asymptotically more efficient alternatives in samples of the size commonly encountered in empirical work.

[10] See Ahn and Schmidt (1995) and Kruiniger (2000).

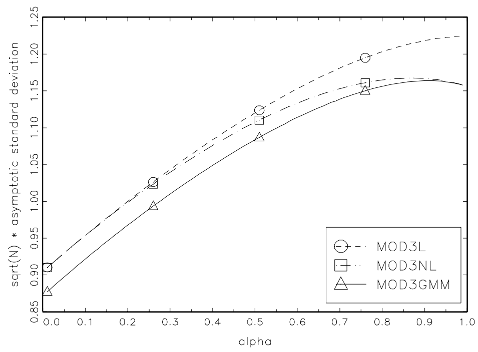

Figure 1. Asymptotic Standard Deviation of $\sqrt{N}\widehat{\alpha}$, two-step.

To conclude this section we investigate how well the finite sample empirical distributions of these estimators for Model 3 are approximated by their asymptotic counterparts.[11] Figures 2a and 2b present p-value plots which compare the actual and nominal sizes of Wald tests of the null hypothesis that α is equal to the true value in our Monte Carlo designs. Results are presented for $N = 500$, $T = 7$ and $\alpha = 0.3$ in Figure 2a, and $\alpha = 0.9$ in Figure 2b. We focus on Wald tests based on the one-step estimators to avoid the problems associated with the estimation of the two-step standard errors. It is clear that for low values of α, the empirical size is well approximated by the nominal size for tests based on all three estimators. In contrast, for $\alpha = 0.9$, the test based on the GMM estimator is considerably oversized, whereas those based on the two projection estimators have much better size properties.

[11] Bond and Windmeijer (2002) study finite sample inference using the GMM estimators for Models 1 and 2 in greater detail.

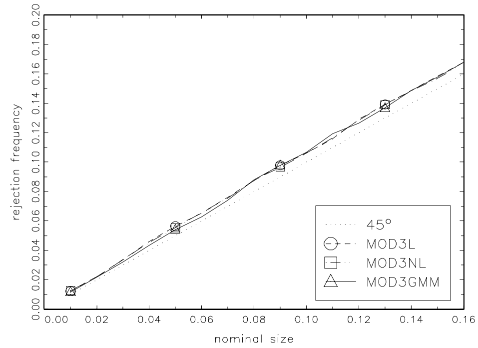

Figure 2a. P-value plot, $\alpha = 0.3$

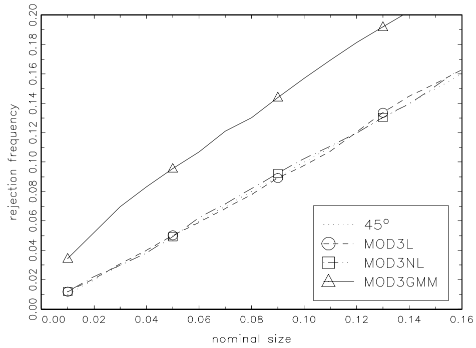

Figure 2b. P-value plot, $\alpha = 0.9$

6. Conclusions

In this paper we have explored a new approach to estimation for autoregressive panel data models, based on projecting the unobserved individual effects on the vector of observations on the lagged dependent variable. The resulting projection estimators coincide with known linear GMM estimators for models where stationarity is not imposed on the initial conditions and for models which satisfy mean stationarity, the differenced GMM and system GMM estimators respectively. We have proposed two new estimators for models which satisfy covariance stationarity, based on the projection approach. One estimator imposes all the restrictions on the projection parameters that are implied by covariance stationarity. The second estimator only imposes simple linear restrictions on the projection parameters. Although the latter is not fully efficient, it is shown in a Monte Carlo study to perform well in terms of small sample bias and precision. This estimator can be implemented straightforwardly using standard regression packages for panel data and in some cases offers a considerable improvement on the system GMM estimator in terms of bias and efficiency.

Acknowledgements

We would like to thank Ekaterini Kyriazidou, two anonymous referees, Min Ahn, Bruno Crépon, Hugo Kruiniger, Jacques Mairesse, Heinz Neudecker and Paul Ruud for helpful comments and suggestions. This work forms part of the research programme of the ESRC Centre for the Microeconomic Analysis of Public Policy at the Institute for Fiscal Studies.

References

[1] Ahn, S.C. and P. Schmidt (1995), Efficient Estimation of Models for Dynamic Panel Data, *Journal of Econometrics*, 68, 5-28.

[2] Ahn, S.C. and P. Schmidt (1997), Efficient Estimation of Dynamic Panel Data Models: Alternative Assumptions and Simplified Estimation, *Journal of Econometrics*, 76, 309-321.

[3] Alonso-Borrego, C. and M. Arellano (1999), Symmetrically Normalized Instrumental-Variable Estimation Using Panel Data, *Journal of Business and Economic Statistics*, 17, 36-49.

[4] 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, 277-98.

[5] Arellano, M. and S. Bond (1998), Dynamic Panel Data Estimation using DPD98 for GAUSS, http://www.ifs.org.uk/staff/steve_b.shtml.

[6] Arellano, M. and O. Bover (1995), Another Look at the Instrumental-Variable Estimation of Error-Components Models, *Journal of Econometrics*, 68, 29-51.

[7] Blundell, R. and S. Bond (1998), Initial Conditions and Moment Restrictions in Dynamic Panel Data Models, *Journal of Econometrics*, 87, 115-143.

[8] Blundell, R., and S. Bond (2000), GMM Estimation with Persistent Panel Data: An Application to Production Functions, *Econometric Reviews* 19, 321-340.

[9] Bond, S., C. Bowsher and F. Windmeijer (2001), Criterion-Based Inference for GMM in Autoregressive Panel Data Models, *Economics Letters* 73, 379-388.

[10] Bond, S., A. Hoeffler and J. Temple (2001), GMM Estimation of Empirical Growth Models, CEPR Discussion Paper No. 3048.

[11] Bond, S. and F. Windmeijer (2002), Finite Sample Inference for GMM Estimators in Linear Panel Data Models, CEMMAP Working Paper No. 04/02, Institute for Fiscal Studies, London.

[12] Chamberlain, G. (1980), Analysis of Covariance with Qualitative Data, *Review of Economic Studies*, 47, 225-238.

[13] Chamberlain, G. (1982), Multivariate Regression Models for Panel Data, *Journal of Econometrics*, 18, 5-46.

[14] Chamberlain, G. (1984), Panel Data, in: Z. Griliches and M.D. Intriligator (eds.), *Handbook of Econometrics*, Volume II, North Holland, Amsterdam, 1247-1318.

[15] Crépon, B. and J. Mairesse (1996), The Chamberlain Approach, in: L. Mátyás and P. Sevestre (eds.), *The Econometrics of Panel Data, A Handbook of Theory with Applications*, 2nd edition, Kluwer Academic Publishers, Dordrecht, 323-391.

[16] Hall, B.H. and C. Cummins (1999), Time Series Processor Version 4.5 User's Guide, TSP International, Palo-Alto, CA.

[17] Holtz-Eakin, D., W. Newey, and H. Rosen (1988), Estimating Vector Autoregressions with Panel Data, *Econometrica,* 56, 1371-1395.

[18] Kruiniger, H. (2000), GMM Estimation of Dynamic Panel Data Models with Persistent Data, Queen Mary, University of London, Dept. of Economics Working Paper No. 428, London.

[19] Ruud, P. (2000), *An Introduction to Classical Econometric Theory*, Oxford University Press.

[20] Sevestre, P. and A. Trognon (1996), Dynamic Linear Models, in: L. Mátyás and P. Sevestre (eds.), *The Econometrics of Panel Data*, *A Handbook of Theory with Applications*, 2nd edition, Kluwer Academic Publishers, Dordrecht, 95-117.

[21] Windmeijer, F. (2000), A Finite Sample Correction for the Variance of Linear Two-Step GMM Estimators, Institute for Fiscal Studies Working Paper No. W00/19, London.

A. Proof that the projection OLS estimator in Model 1 coincides with the within groups estimator

Stack the observations in such a way that the projection model (4.2) can be written as

$$y = \alpha y_{-1} + (\iota \otimes Y_{-1})\,\delta + \tilde{\eta} + v$$

where $Y_{-1} = \left[\underline{y}_1^{T-1}, ..., \underline{y}_N^{T-1}\right]'$ is an $N \times (T-1)$ matrix, $y_{-1} = vec\,(Y_{-1})$ and ι is a vector of ones of order $T-1$. The projection of y_{-1} on $(\iota \otimes Y_{-1})$ is given by

$$\frac{1}{T-1}\left(\iota\iota' \otimes Y_{-1}\left(Y_{-1}'Y_{-1}\right)^{-1}Y_{-1}'\right)y_{-1}.$$

But

$$\begin{aligned}
\frac{1}{T-1}\left(\iota\iota' \otimes Y_{-1}\left(Y_{-1}'Y_{-1}\right)^{-1}Y_{-1}'\right)y_{-1} &= \frac{1}{T-1}\left(\iota\iota' \otimes Y_{-1}\left(Y_{-1}'Y_{-1}\right)^{-1}Y_{-1}'\right)vec\,(Y_{-1}) \\
&= \frac{1}{T-1}vec\,(Y_{-1}\iota\iota') \\
&= \frac{1}{T-1}\left(\iota\iota' \otimes I_n\right)vec\,(Y_{-1}) = \frac{1}{T-1}\left(\iota\iota' \otimes I_n\right)y_{-1}.
\end{aligned}$$

Therefore using standard results for partitioned regression the OLS estimator of α in the projection model is numerically identical to the within groups estimator.

Note that this result does not depend on the fact that the regressor considered here is the lagged dependent variable. In the more general specification with a vector of regressors x_{it}

$$y_{it} = \beta' x_{it} + \eta_i + v_{it},$$

with associated projection model

$$y_{it} = \beta' x_{it} + \sum_{j=1}^{T}\delta_{xj}' x_{ij} + \tilde{\eta}_i + v_{it}, \qquad \text{(A.1)}$$

a similar argument shows that the OLS estimator for β in (A.1) is equivalent to the within groups estimator.

B. Proof that the projection 2SLS estimator in Model 1 coincides with the one-step differenced GMM estimator using the optimal weight matrix when the v_{it} are homoskedastic

A one-step differenced GMM estimator can be obtained by 2SLS as

$$
\begin{aligned}
& \left(y_{-1}' D Z_d \left(Z_d' D' D Z_D\right)^{-1} Z_d' D' y_{-1}\right)^{-1} y_{-1}' D Z_d \left(Z_d' D' D Z_d\right)^{-1} Z_d' D' y \\
= & \left(y_{-1}' P_{Z_d^*} y_{-1}\right)^{-1} y_{-1}' P_{Z_d^*} y
\end{aligned}
$$

where

$$
Z_d = \begin{bmatrix} y_1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & y_1 & y_2 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \ddots & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & y_1 & \dots & y_{T-2} \end{bmatrix}; \quad D = \begin{bmatrix} -I_N & 0 & 0 & 0 \\ I_N & -I_N & 0 & 0 \\ 0 & I_N & \ddots & 0 \\ 0 & 0 & \ddots & -I_N \\ 0 & 0 & 0 & I_N \end{bmatrix};
$$

$Z_d^* = D Z_d$ and $P_A = A\left(A'A\right)^{-1} A'$ for some matrix A. Note that $D'D = H \otimes I_N$ and so $\left(Z_d' D' D Z_d\right)^{-1}$ is the optimal weight matrix when the v_{it} are homoskedastic.

The 2SLS projection estimator is OLS in the model

$$
y = \alpha y_{-1} + \widehat{X}\delta + \tilde{\eta} + v.
$$

where

$$
\widehat{X} = P_{Z_p}\left(\iota \otimes Y_{-1}\right)
$$

with

$$
Z_p = \begin{bmatrix} Z_d & 0 \\ 0 & Y_{-1} \end{bmatrix}
$$

The 2SLS projection estimator for α is given by

$$
\left(y_{-1}' M_{\widehat{X}} y_{-1}\right)^{-1} y_{-1}' M_{\widehat{X}} y
$$

where

$$
M_{\widehat{X}} = I_{N(T-1)} - P_{\widehat{X}}.
$$

As

$$\left(P_{Z_d^*} - M_{\widehat{X}}\right) y_{-1} = 0,$$

the two estimators are the same.

Proof:

The prediction of X is given by

$$\widehat{X} = \begin{bmatrix} y_1 & P_{y_1} y_2 & P_{y_1} y_3 & P_{y_1} y_4 & & & P_{y_1} y_{T-1} \\ y_1 & y_2 & P_{Y_2} y_3 & P_{Y_2} y_4 & & & P_{Y_2} y_{T-1} \\ y_1 & y_2 & y_3 & P_{Y_3} y_4 & & & P_{Y_3} y_{T-1} \\ y_1 & y_2 & y_3 & y_4 & & & P_{Y_4} y_{T-1} \\ & & & & \ddots & & \vdots \\ & & & & & \ddots & P_{Y_{T-2}} y_{T-1} \\ y_1 & y_2 & y_3 & y_4 & \cdots & y_{T-2} & y_{T-1} \end{bmatrix},$$

where $Y_q = [y_1, ..., y_q]$, and therefore

$$Z_d^{*\prime} \widehat{X} = 0.$$

From this it follows that

$$\left(P_{Z_d^*} + P_{\widehat{X}}\right) \widehat{X} = \widehat{X} \quad \text{(B.1)}$$

$$\left(P_{Z_d^*} + P_{\widehat{X}}\right) Z_d^* = Z_d^*. \quad \text{(B.2)}$$

Denote $Q = P_{Z_d^*} + P_{\widehat{X}}$, then it follows from (B.1)

$$Q\left(\iota_{T-2} \otimes y_1\right) = \left(\iota_{T-2} \otimes y_1\right), \quad \text{(B.3)}$$

where ι_{T-2} is a vector of ones of order $T-2$, and from (B.2)

$$QD\left(I_{T-2} \otimes y_1\right) = D\left(I_{T-2} \otimes y_1\right) \quad \text{(B.4)}$$

Combining (B.3) with (B.4), it follows that

$$Q\left(I_{T-2} \otimes y_1\right) = \left(I_{T-2} \otimes y_1\right). \quad \text{(B.5)}$$

From the second column of $\widehat{X}$, we get

$$Q\begin{bmatrix} P_{y_1}y_2 \\ \iota_{T-3}\otimes y_2 \end{bmatrix} = \begin{bmatrix} P_{y_1}y_2 \\ \iota_{T-3}\otimes y_2 \end{bmatrix}$$

but from (B.5) we already know that

$$Q\begin{bmatrix} P_{y_1}y_2 \\ 0 \end{bmatrix} = \begin{bmatrix} P_{y_1}y_2 \\ 0 \end{bmatrix}$$

and therefore

$$Q\begin{bmatrix} 0 \\ \iota_{T-3}\otimes y_2 \end{bmatrix} = \begin{bmatrix} 0 \\ \iota_{T-3}\otimes y_2 \end{bmatrix}. \tag{B.6}$$

From (B.2) we have that

$$QD\begin{bmatrix} 0 \\ I_{T-3}\otimes y_2 \end{bmatrix} = D\begin{bmatrix} 0 \\ I_{T-3}\otimes y_2 \end{bmatrix} \tag{B.7}$$

and again, combining (B.6) and (B.7) we get

$$Q\begin{bmatrix} 0 \\ I_{T-3}\otimes y_2 \end{bmatrix} = \begin{bmatrix} 0 \\ I_{T-3}\otimes y_2 \end{bmatrix}.$$

Repeating these steps, we get the result that

$$\left(P_{Z_d^*} + P_{\widehat{X}}\right) Z_p = Z_p,$$

or

$$\left(P_{Z_d^*} - M_{\widehat{X}}\right) Z_p = 0,$$

from which it follows that

$$\left(P_{Z_d^*} - M_{\widehat{X}}\right) y_{-1} = 0.$$

C. Proof that the projection 2SLS estimator in Model 2 coincides with the one-step system GMM estimator using the efficient weight matrix when the v_{it} are homoskedastic and $\boldsymbol{\eta}_i = \mathbf{0}$ for $i = 1, ..., N$

The restricted model is

$$y = \alpha y_{-1} + \left(\iota \otimes Y_{-1}\left(\frac{1}{N}Y_{-1}'Y_{-1}\right)^{-1}\iota\right)\kappa + \widetilde{\eta} + v.$$

The one-step system estimator can be obtained as

$$\begin{aligned} & \left(y_{-1}' S Z_s \left(Z_s' S' S Z_s\right)^{-1} Z_s' S' y_{-1}\right)^{-1} y_{-1}' S Z_s \left(Z_s' S' S Z_s\right)^{-1} Z_s' S' y \\ = & \left(y_{-1}' P_{Z_s^*} y_{-1}\right)^{-1} y_{-1}' P_{Z_s^*} y \end{aligned}$$

where

$$Z_s = \begin{bmatrix} Z_d & 0 & 0 & 0 & 0 \\ 0 & \Delta y_2 & 0 & 0 & 0 \\ 0 & 0 & \Delta y_3 & 0 & 0 \\ 0 & 0 & 0 & \ddots & 0 \\ 0 & 0 & 0 & 0 & \Delta y_{T-1} \end{bmatrix}; \; S = \begin{bmatrix} -I_N & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ I_N & -I_N & 0 & 0 & I_N & 0 & 0 & 0 \\ 0 & I_N & \ddots & 0 & 0 & I_N & 0 & 0 \\ 0 & 0 & \ddots & -I_N & 0 & 0 & \ddots & 0 \\ 0 & 0 & 0 & I_N & 0 & 0 & 0 & I_N \end{bmatrix}$$

and $Z_s^* = S Z_s$.

Note that $S'S = A \otimes I_N$ and so $\left(Z_s' S' S Z_s\right)^{-1}$ is the optimal weight matrix when the v_{it} are homoskedastic and $\eta_i = 0$ for $i = 1, ..., N$.

The 2SLS projection estimator is OLS in the model

$$\widehat{y} = \alpha y_{-1} + \widehat{x}_r \kappa + \widetilde{\eta} + v.$$

where

$$\widehat{x}_r = P_{Z_p} \left(\iota \otimes Y_{-1} \left(\frac{1}{N} Y_{-1}' Y_{-1} \right) \iota \right)$$

The 2SLS projection estimator for α is given by

$$\left(y_{-1}' M_{\widehat{x}_r} y_{-1}\right)^{-1} y_{-1}' M_{\widehat{x}_r} y$$

where

$$M_{\widehat{x}_r} = I_{N(T-1)} - P_{\widehat{x}_r}.$$

As

$$\left(P_{Z_s^*} - M_{\widehat{x}_r}\right) y_{-1} = 0,$$

which can be shown analogously to the proof in the previous section, the two estimators are the same.

D. A model with an additional regressor

The model we consider here is

$$y_{it} = \alpha y_{i,t-1} + \beta x_{it} + \eta_i + v_{it},$$

for $t = 2, ..., T$. The projection model is specified either as

$$y_{it} = \alpha y_{i,t-1} + \beta x_{it} + \delta_y' \underline{y}_i^{T-1} + \delta_x' \underline{x}_i^T + \widetilde{\eta}_i + v_{it}, \tag{D.1}$$

or

$$y_{it} = \alpha y_{i,t-1} + \beta x_{it} + \delta_y' \underline{y}_i^{T-1} + \delta_x' \underline{x}_i^{T-1} + \widetilde{\eta}_i + v_{it}, \tag{D.2}$$

where $\underline{y}_i^{T-1} = (y_{i1}, y_{i2}, ..., y_{i,T-1})'$, $\underline{x}_i^T = (x_{i1}, x_{i2}, ..., x_{iT})'$ and $\underline{x}_i^{T-1} = (x_{i1}, x_{i2}, ..., x_{i,T-1})'$.

Which of these projections is appropriate depends on the correlation between x_{it} and v_{is} and hence on the values of the x_{it} series that are used as instruments for estimating these projection models by 2SLS or GMM. For example, if x_{it} is endogenous such that $E(v_{it}x_{it}) \neq 0$, but $E(v_{it}x_{is}) = 0$ for $s < t$, then x_{iT} is not available as an instrument in the projection model for period T (or for any earlier periods). In this case, only $x_{i1}, ..., x_{i,T-1}$ are used as instruments and the projection specification (D.2) is appropriate. Notice that in this case, x_{it} and y_{it} are treated symmetrically. On the other hand, if x_{it} is predetermined such that $E(v_{it}x_{is}) = 0$ for $s \leq t$ (or strictly exogenous such that $E(v_{it}x_{is}) = 0$ for all s, t) then x_{iT} will be used as an instrument in the projection model for period T (or for all periods in the strictly exogenous case) and therefore the projection specification (D.1) is required to ensure that $\widetilde{\eta}_i$ is orthogonal to the complete set of instruments. A proof similar to that in section B can be used to establish that the 2SLS estimators for these projection models coincide with the corresponding one-step differenced GMM estimators, described in Arellano and Bond (1991), when the optimal weight matrix under homoskedasticity of v_{it} is used.[12]

[12] Implicitly, equation (A.1) defines a third projection specification for this model, where η_i is

To impose mean stationarity on the projection model, we can proceed as follows. Assuming that both x_{it} and y_{it} are mean stationary we define[13]

$$\begin{aligned} E\left(x_{it}\eta_i\right) &= \kappa_x \\ E\left(y_{it}\eta_i\right) &= \kappa_y, \end{aligned}$$

for $t = 1, ..., T$. The projection parameters δ_y and δ_x are given by

$$\left(\begin{array}{c} \delta_y \\ \delta_x \end{array}\right) = E\left[\begin{array}{cc} \underline{y}_i^{T-1}\left(\underline{y}_i^{T-1}\right)' & \underline{y}_i^{T-1}\left(\underline{x}_i^{R}\right)' \\ \underline{x}_i^{R}\left(\underline{y}_i^{T-1}\right)' & \underline{x}_i^{R}\left(\underline{x}_i^{R}\right)' \end{array}\right]^{-1} E\left(\begin{array}{c} \underline{y}_i^{T-1}\eta_i \\ \underline{x}_i^{R}\eta_i \end{array}\right),$$

where $R = T$ or $R = T - 1$, depending on whether the projection specification (D.1) or (D.2) is used. Let

$$\left[\begin{array}{cc} E & F \\ F' & G \end{array}\right] = N\left[\begin{array}{cc} Y_{-1}'Y_{-1} & Y_{-1}'X \\ X'Y_{-1} & X'X \end{array}\right]^{-1},$$

$Y_{-1} = \left[\underline{y}_1^{T-1}, ..., \underline{y}_N^{T-1}\right]'$ and $X = \left[\underline{x}_1^{R}, ..., \underline{x}_N^{R}\right]'$, then the projection model which imposes the mean-stationarity restrictions is

$$y_{it} = \alpha y_{i,t-1} + \beta x_{it} + \kappa_y\left(\left(\underline{y}_i^{T-1}\right)' E\iota + \underline{x}_i^{R\prime}F'\iota\right) + \kappa_x\left(\left(\underline{y}_i^{T-1}\right)' F\iota + \underline{x}_i^{R\prime}G\iota\right) + \widetilde{\eta}_i + v_{it}.$$

The parameters α, β, κ_y, and κ_x are estimated by 2SLS or GMM using the appropriate instruments, which will again depend on the assumed correlation between x_{it} and v_{is}, as discussed above. A proof similar to that in section C can be used to establish that the 2SLS estimators for these projection models coincide with the corresponding one-step system GMM estimators, when the optimal weight matrix under homoskedasticity of v_{it} and $\eta_i = 0$ for $i = 1, ..., N$ is used.

---

projected on $\underline{y}_i^{T-1}$ and $(x_{i2}, ..., x_{iT})$. As shown in section A, OLS on this projection specification coincides with the within groups estimator. It can also be shown that OLS on (D.1) coincides with within groups, but the same is not true for OLS on (D.2).

[13] If for example $x_{it} = \gamma x_{i,t-1} + \xi_i + \varepsilon_{it}$, with η_i and ξ_i homoskedastic across individuals, $E\left(\eta_i^2\right) = \sigma_\eta^2$, $E\left(\xi_i^2\right) = \sigma_\xi^2$ and $E\left(\eta_i\xi_i\right) = \sigma_{\eta\xi}$, then $\kappa_x = \frac{\sigma_{\eta\xi}}{1-\gamma}$ and $\kappa_y = \frac{\sigma_\eta^2}{1-\alpha} + \frac{\beta\sigma_{\eta\xi}}{(1-\alpha)(1-\gamma)}$.