Initial Conditions and Moment Restrictions in Dynamic Panel Data Models

Richard Blundell*
University College London and Institute for Fiscal Studies

Stephen Bond†
Nuffield College and Institute for Fiscal Studies

October 1995

*Department of Economics, University College London, Gower Street, London, WC1E 6BT.
†Nuffield College, Oxford, OX1 1NF.

Summary

In this paper we consider estimation of the autoregressive error components model $y_{it} = \alpha y_{i,t-1} + \eta_i + v_{it}$. When the autoregressive parameter α is moderately large and the number of time series observations is moderately small, the usual Generalised Methods of Moments (GMM) estimator obtained after first differencing has been found to be poorly behaved. Here we consider alternative linear estimators that are designed to improve the properties of the standard first-differenced GMM estimator.

We consider two approaches to estimation. The first approach extends the model by adding the observed initial values as an extra regressor. This allows consistent estimates to be obtained by error-components GLS. This estimator is shown to be equivalent to the optimal GMM estimator for the normal homoskedastic error components model. The second approach considers a mild restriction on the initial condition process under which lagged Δy_{it} can be used to construct linear moment conditions in the levels equations. The complete set of moment conditions can then be exploited by a linear GMM estimator in a system of first-differenced and levels equations, rendering the non-linear moment conditions redundant for estimation. This estimator is strictly more efficient than non-linear GMM when the additional restriction is valid. Monte Carlo simulations are reported which demonstrate the dramatic improvement in performance of the proposed estimators compared to the usual first-differenced GMM estimator, especially for high values of the autoregressive parameter α.

Acknowledgements

We are grateful to Seung Ahn, Manuel Arellano, Andrew Chesher, Bronwyn Hall, Jacques Mairesse, Whitney Newey, Neil Shephard, Richard Smith, Richard Spady, Frank Windmeijer and seminar participants at INSEE, Manchester, MIT and Nuffield for helpful comments. Financial support from the ESRC is gratefully acknowledged. This research is part of the programme of research at the ESRC Centre for the Micro-Economic Analysis of Fiscal Policy at IFS.

1. Introduction

In dynamic panel data models where the autoregressive parameter is moderately large and the number of time series observations is moderately small, the linear Generalised Methods of Moments (GMM) estimator obtained after first differencing has been found to be poorly behaved. Lagged levels of the series provide weak instruments for first differences in this case. Non-linear moment conditions that could be expected to improve this behaviour are available for these models but have not been widely adopted in practice. Here we consider alternative estimators that are designed to improve the properties of the standard first-differenced GMM estimator and are conveniently linear. We first show that the error components GLS estimation of a model that conditions on the initial values provides a consistent estimator, asymptotically equivalent to the non-linear GMM estimator in the homoskedastic normal autoregressive model. A Monte Carlo analysis also suggests that it has good finite sample properties. We then show that relatively mild restrictions on the initial conditions process can be used to both simplify the non-linear moment conditions and increase precision. The resulting linear estimator uses lagged *differences* of y_{it} as instruments for equations in levels, in addition to lagged *levels* of y_{it} as instruments for equations in first differences. Monte-Carlo simulations of this linear system estimator and asymptotic variance calculations show that this offers dramatic efficiency gains in the situations where the basic first-differenced GMM estimator performs poorly.

The gain in precision that results from exploiting the initial condition information is shown to increase for higher values of the autoregressive parameter and as the number of time series observations gets smaller. For short panels we find both a large downward bias and very low precision for the standard first-differenced estimator when the autoregressive parameter is high. The initial condition information exploited in our system estimators not only greatly improves the precision but also greatly reduces the finite sample bias in this case. We also observe that whilst the finite sample bias of the differenced estimator is generally downwards (in the direction of OLS in first differences), the much smaller finite sample bias of the system estimator is generally upwards (in the direction of OLS levels).

The layout of the paper is as follows. In Section 2 we briefly review the standard moment conditions for the autoregressive error components model in the framework of Anderson and Hsiao (1982), Holtz-Eakin, Newey and Rosen (1988), Arellano and Bond (1991) and Ahn and Schmidt (1995). In Section 3 we consider the conditional GLS estimator, introduced by Blundell and Smith (1991), in which initial conditions are explicitly added to the model. In Section 4 we consider restrictions on the initial condition process that render lagged values of Δy_{it} valid as instruments for the levels equations. The resulting estimator is equivalent to that recently proposed by Arellano and Bover (1995). Section 5 presents the results of Monte Carlo simulations which highlight the potential importance of exploiting the extra moment restrictions relating to the properties of the initial condition process for the efficiency of the AR coefficient estimators. Section 6 discusses the extensions to models with strictly exogenous and predetermined regressors, and concludes.

2. The Moment Conditions

We consider an AR(1) model with unobserved individual-specific effects

$$y_{it} = \alpha y_{i,t-1} + \eta_i + v_{it} \tag{2.1}$$

for $i = 1, ..., N$ and $t = 2, ..., T$, where $\eta_i + v_{it} = u_{it}$ is the usual "fixed effects" decomposition of the error term; N is large and T is fixed.[1] Since our focus is on the role of initial conditions we will assume, for most of our discussion, that η_i and v_{it} have the familiar error-components structure in which

$$E(\eta_i) = 0 \ , \ E(v_{it}) = 0 \ , \ E(v_{it}\eta_i) = 0 \text{ for } i = 1, ..., N \text{ and } t = 2, ..., T \tag{2.2}$$

and

$$E(v_{it}v_{is}) = 0 \text{ for } i = 1, ..., N \text{ and } \forall \ t \neq s. \tag{2.3}$$

In addition there is the standard assumption concerning the initial conditions y_{i1}(see Ahn and Schmidt (1995), for example)

$$E(y_{i1}v_{it}) = 0 \text{ for } i = 1, ..., N \text{ and } t = 2, ..., T. \tag{2.4}$$

Conditions (2.2), (2.3) and (2.4) imply moment restrictions that are sufficient to (identify and) estimate α. However, we may also wish to incorporate the homoskedasticity restrictions

$$E(v_{it}^2) = \sigma_v^2, \ E(\eta_i^2) = \sigma_\eta^2 \text{ for } i = 1, ..., N \text{ and } t = 2, ..., T. \tag{2.5}$$

[1] All of the estimators discussed below and their properties extend in an obvious fashion to higher order autoregressive models. In the concluding section we also consider extensions to models that include regressors.

so that

$$E(\mathbf{u}_i\mathbf{u}_i') = \mathbf{\Omega} = \sigma_\eta^2\mathbf{J}_{T-1} + \sigma_v^2\mathbf{I}_{T-1}.$$

where $\mathbf{u}_i' = (u_{i2}, ..., u_{iT})$ and $\mathbf{J}_{T-1}$ is the (T-1)x(T-1) unit matrix and $\mathbf{I}_{T-1}$ the (T-1)x(T-1) identity matrix.

Finally, in our discussion of maximum likelihood estimation, we will make use of the further conditional normality assumption

$$\mathbf{u}_i|y_{i1} \sim N(\ \mathbf{0},\ \mathbf{\Omega}\ ) \text{ for all } i = 1, ...N. \tag{2.6}$$

In the absence of any further restrictions on the process generating the initial conditions, the autoregressive error components model (2.1) - (2.5) implies the following orthogonality conditions

$$E(y_{i,t-s}\Delta v_{it}) = 0; \text{ for } t = 3, ..., T \text{ and } s \geq 2 \tag{2.7}$$

$$E(y_{i,t-2}\Delta v_{i,t-1} - y_{i,t-1}\Delta v_{it}) = 0; \text{ for } t = 4, ..., T \tag{2.8}$$

which are linear in the α parameter. Although these moment conditions can be expressed in alternative ways, the organisation of (2.7) and (2.8) will prove useful in what follows. The 0.5(T-1)(T-2) orthogonality conditions in (2.7) simply reflect the assumed absence of serial correlation in the time varying disturbances v_{it}, together with the restriction (2.4). The T-3 conditions in (2.8) depend on the homoskedasticity through time of v_{it}.

The moment restrictions in (2.7) and (2.8) can be expressed more compactly as

$$E(\mathbf{Z}_i'\overline{\mathbf{u}}_i) = \mathbf{0}$$

where $\mathbf{Z}_i$ is the (T-2)xm matrix given by (omitting the i subscripts)

$$\mathbf{Z}_i = \begin{bmatrix} y_1 & y_2 & 0 & 0 & 0 & \ldots & 0 & 0 & \ldots & 0 \\ 0 & -y_3 & y_1 & y_2 & y_3 & \ldots & 0 & 0 & \ldots & 0 \\ . & . & . & . & . & \ldots & . & . & \ldots & . \\ 0 & 0 & 0 & 0 & 0 & \ldots & -y_{T-1} & y_1 & \ldots & y_{T-2} \end{bmatrix},$$

$\overline{\mathbf{u}}_i$ is the (T-2) vector $(\Delta v_{i3,}\Delta v_{i4}, ..., \Delta v_{iT})'$, and $m = 0.5(T-1)(T-2) + (T-3)$.

The Generalised Method of Moments estimator based on these moment conditions minimises the quadratic distance $(\overline{\mathbf{u}}'\mathbf{Z}\mathbf{A}_N\mathbf{Z}'\overline{\mathbf{u}})$ for some metric $\mathbf{A}_N$, where $\mathbf{Z}'$ is the mxN(T-2) matrix $(\mathbf{Z}_1', \mathbf{Z}_2', ..., \mathbf{Z}_N')$ and $\overline{\mathbf{u}}$ is the N(T-2) vector $(\overline{\mathbf{u}}_1', \overline{\mathbf{u}}_2', ..., \overline{\mathbf{u}}_N')'$. This gives the GMM estimator for α as

$$\widehat{\alpha} = (\overline{\mathbf{y}}_{-1}'\mathbf{Z}\mathbf{A}_N\mathbf{Z}'\overline{\mathbf{y}}_{-1})^{-1}\overline{\mathbf{y}}_{-1}'\mathbf{Z}\mathbf{A}_N\mathbf{Z}'\overline{\mathbf{y}}.$$

where $\overline{\mathbf{y}}_i$ is the (T-2) vector $(\Delta y_{i3,}\Delta y_{i4}, ..., \Delta y_{iT})'$.

Alternative choices for the weights $\mathbf{A}_N$ give rise to a set of GMM estimators based on the moment conditions in (2.7) and (2.8), all of which are consistent for large N and finite T, but which differ in their asymptotic efficiency. In general the optimal weights are given by

$$\mathbf{A}_N = (N^{-1}\sum_{i=1}^{N}\mathbf{Z}_i'\widehat{\overline{\mathbf{u}}}_i\widehat{\overline{\mathbf{u}}}_i'\mathbf{Z}_i)^{-1}$$

where $\widehat{\overline{\mathbf{u}}}_i$ are residuals from an initial consistent estimator. We refer to this as the *two-step* GMM estimator[2]. In the absence of any additional knowledge about the process for the initial conditions, this estimator is asymptotically efficient in

[2]As a choice of $\mathbf{A}_N$ to yield the initial consistent estimator we consider

$$\mathbf{A}_N = (N^{-1}\sum_{i=1}^{N}\mathbf{Z}_i'\mathbf{H}\mathbf{Z}_i)^{-1}$$

the class of estimators based on the linear moment conditions (2.7) and (2.8) (see Hansen (1982)).

In a recent contribution, Crepon, Kramarz and Trognon (1993) have noted that the autoregressive error components model implies a further (T-1) linear moment conditions which are given by

$$E(u_{it}) = 0 \text{ for } t = 2, ..., T. \tag{2.9}$$

This implies that the (T-1) time dummies could be used as instruments for the equations in levels.[3]

Whether these extra moment conditions (2.9) will be helpful in estimating α depends crucially on the nature of the initial conditions. In the stationary model we also have

$$E(y_{it}) = 0 \text{ for } t = 1, ..., T.$$

In this case the use of time dummies as instruments would not be informative. However, for other "start-up" processes this need not be the case (for example, if $y_{i1} = k_i$ with $E(k_i) \neq 0$).

---

where $\mathbf{H}$ is the (T-2)x(T-2) matrix given by

$$\mathbf{H} = \begin{pmatrix} 2 & -1 & 0 & ... & 0 \\ -1 & 2 & -1 & ... & 0 \\ 0 & -1 & 2 & ... & 0 \\ ... & ... & ... & ... & . \\ 0 & 0 & 0 & ... & 2 \end{pmatrix}$$

which can be calculated in one step. Note that when the v_{it} are i.i.d., the one-step and two-step estimators are asymptotically equivalent in this model.

[3]It should be noted that time effects are typically included in dynamic panel data models. Time dummies in the differenced equations would therefore be automatically included in the instrument set, (2.9) then implies only one extra moment condition, which could be exploited in a system of first-differenced and levels equations as will be considered in section 4.

Finally, a number of authors have suggested using the additional T-3 non-linear moment conditions (see Ahn and Schmidt, 1995, eq(4))

$$E(u_{it}\Delta u_{i,t-1}) = 0; \text{ for } t = 4, 5, ..., T. \tag{2.10}$$

which could be expected to improve efficiency and clearly are implied by (2.3) and (2.4). These conditions relate directly to the absence of serial correlation in v_{it} and do not require homoskedasticity. Under homoskedasticity Ahn and Schmidt (1995, eq(11b)) show the existence of an additional non-linear moment condition

$$E(\overline{u}_i \Delta u_{i3}) = 0 \text{ where } \overline{u}_i = \frac{1}{T-1}\sum_{t=2}^{T} u_{it}. \tag{2.11}$$

These non-linear moments are likely to be particularly informative in the case where α is close to unity: estimators which rely on lagged levels as instruments for current differences are likely to perform poorly when the series are close to random walks. Here we consider two linear estimators which share this advantage: a conditional GLS estimator which is equivalent to non-linear GMM under normality and homoskedasticity; and a linear GMM estimator which is more efficient but relies on a restriction on the initial condition process.

3. The Conditional GLS Estimator

In the autoregressive error components model, the homoskedasticity assumption (2.5) can be used to derive a consistent *conditional* GLS (CGLS) estimator by including y_{i1} in each of the T-1 levels equations and then applying the standard error-components GLS estimator. Under normality, this is the conditional ML estimator (CMLE) proposed by Blundell and Smith (1991) and below it is shown to be asymptotically equivalent to the optimally weighted non-linear GMM estimator.

The presence of η_i not only implies

$$E(\mathbf{y}'_{i,-1}\mathbf{u}_i) \neq \mathbf{0} \tag{3.1}$$

where $\mathbf{y}'_{i,-1} = (y_{i1}, ..., y_{i,T-1})$, but also

$$E(\mathbf{y}'_{i,-1}\boldsymbol{\Omega}^{-1}\mathbf{u}_i) \neq \mathbf{0}. \tag{3.2}$$

As a result the standard OLS, Within Groups and GLS estimators are all inconsistent (see Hsiao (1986), Nickell (1981) and Sevestre and Trognon (1985), respectively). The inconsistency of the GLS estimator can be seen to depend on $E(y_{i1}\eta_i) \neq 0$. That is, premultiplying the T-1 system (2.1) by $\boldsymbol{\Omega}^{-1/2}$ introduces y_{i1} into all T-1 equations of the transformed system, inducing a correlation with η_i. The aim of the conditional GLS estimator is to eliminate this correlation.

Given (2.4), we have that $E(y_{i1}u_{it}) = E(y_{i1}\eta_i)$ for $t = 2, ..., T$. The linear projection of η_i on y_{i1} then defines

$$proj(\eta_i|y_{i1}) = \tau y_{i1} \text{ for } t = 2, ..., T \text{ and } i = 1, ..., N. \tag{3.3}$$

Including y_{i1} in (2.1) gives

$$y_{it} = \alpha y_{i,t-1} + \tau y_{i1} + \tilde{\eta}_i + v_{it} \text{ for } t = 2, ..., T \text{ and } i = 1, ..., N, \tag{3.4}$$

in which $\tilde{\eta}_i = \eta_i - \tau y_{i1}$ and $E(y_{i1}\tilde{\eta}_i) = 0$. The conditional model (3.4) retains an error components structure so that, defining $\tilde{\boldsymbol{\Omega}}$ to be the (T-1)x(T-1) error covariance matrix for the errors $\widetilde{u}_{it} = \widetilde{\eta}_i + v_{it}$ in the stacked conditional model, we have the following moment condition

$$E(\mathbf{y}_{i,-1}' \bar{\mathbf{\Omega}}_i^{-1} \widetilde{\mathbf{u}}_i) = \mathbf{0}. \tag{3.5}$$

Error-components GLS will therefore be consistent for α in the conditional model (3.4)[4].

Ahn and Schmidt (1995) show that the optimally weighted GMM estimator that uses moment conditions (2.7) to (2.11) is asymptotically efficient among all estimators that use second moment information. Under normality this is equivalent to the conditional maximum likelihood estimator. Otherwise GMM is strictly more efficient (see Chamberlain (1987)). The equivalence between the conditional MLE (CGLS) and optimally weighted GMM is given in the following proposition:

Proposition 3.1. *For the normal linear autoregressive error components model (2.1) - (2.6), the Conditional GLS estimator is asymptotically equivalent to the optimally weighted GMM estimator.*

Proof. See Appendix A.

4. Non-linear GMM and the Initial Condition Process

In this section we consider an additional but in many cases relatively mild restriction on the initial conditions process which allows the use of additional linear

[4]Note that it remains the case that $E(\mathbf{y}_{i,-1}' \widetilde{\mathbf{u}}_i) \neq \mathbf{0}$. The orthogonality condition (3.5) is an example of a GLS transformation which is necessary for consistency and not just efficiency. As a result, although consistency does not require normality, it does require imposing the correct restrictions on $\bar{\mathbf{\Omega}}$. As indicated in Prucha (1987), this is a situation when the asymptotic covariance matrix for the *feasible* conditional GLS of α depends on the distribution of the estimated $\bar{\mathbf{\Omega}}$. Bootstrap methods would, however, be applicable in this parametric autoregressive model as described in Horowitz (1995).

moment conditions for the levels equations in the GMM framework. This allows the use of lagged differences of y_{it} as instruments in the levels equations. These additional moment conditions are likely to be important in practice when α is close to unity, since lagged values in the differenced equations will be weak instruments in this case. Very conveniently, these linear moment conditions also imply the non-linear conditions (2.10), from which it follows that the optimal GMM estimator can be implemented as a linear GMM estimator under this restriction.[5]

4.1. An additional moment restriction

In contrast to the non-linear moment restrictions (2.10), we consider the following T-3 linear moment conditions

$$E(u_{it}\Delta y_{i,t-1}) = 0; \text{ for } t = 4, 5, ..., T. \tag{4.1}$$

The use of lagged differences as possible instruments for equations in levels was proposed by Arellano and Bover (1995). Clearly (4.1) does not imply (2.10). However, notice that since Δy_{i2} is observed, there is an additional restriction available, namely

$$E(u_{i3}\Delta y_{i2}) = 0. \tag{4.2}$$

Note also that the validity of this extra moment condition depends on a restriction on the initial condition process generating y_{i1}. Condition (4.2) has two important implications. First, combining (4.2) with the model for periods 2,...,T set out in

[5]In this section we focus on moment conditions that remain valid under heteroskedasticity. Ahn and Schmidt (1995, equation (12b)) have shown that the homoskedasticity restrictions (2.8) and (2.11) can be replaced by a set of T-2 linear moment conditions under the restriction (4.2) specified below.

(2.1)-(2.4) implies the validity of the linear moment restrictions in (4.1)[6]. Second, combining (4.1) and (4.2) implies the non-linear restrictions in (2.10), and renders these redundant for estimation.

To examine the conditions under which (4.2) will hold, we consider separately the cases where $\alpha < 1$ and $\alpha = 1$.

4.1.1. $\alpha < 1$

For $\alpha < 1$ we write y_{i1} as

$$y_{i1} = \frac{\eta_i}{1-\alpha} + u_{i1}. \tag{4.3}$$

The model specifies a convergent level for y_{it} from $t = 2$ onwards for each individual, and u_{i1} is simply the initial deviation from this convergent level. Necessary conditions for (4.2) are then

$$E(u_{i1}\eta_i) = E(u_{i1}v_{i3}) = 0 \text{ for } i = 1, ..., N. \tag{4.4}$$

The key requirement is therefore that the initial *deviations* from $\eta_i/(1-\alpha)$ are not correlated with the *level* of $\eta_i/(1-\alpha)$ itself.

This condition is clearly satisfied in the fully stationary model, where u_{i1} will be the infinite weighted sum $\sum_{s=-1}^{\infty}\left(\alpha^{s+1}v_{i,-s}\right)$. However, stationarity is not necessary for the validity of the extra linear moment conditions in (4.2) and (4.1). Condition (4.4) imposes no restriction on the variance of u_{i1}, and any entry period

[6]Since the model implies

$$\Delta y_{it} = \alpha^{t-2}\Delta y_{i2} + \sum_{s=0}^{t-3}\alpha^s \Delta v_{i,t-s} \text{ for } t = 3, 4, ..., T.$$

"disequilibrium" from $\eta_i/(1-\alpha)$ which is randomly distributed across agents will preserve condition (4.2). Other departures from stationarity such as $y_{i1} = k \; \forall \; i$ or $y_{i1} \sim i.i.d.(0, \sigma^2_{y_1})$ will violate (4.4) however, so this requirement is not trivial.

4.1.2. $\alpha = 1$

In the case where $\alpha = 1$, condition (4.2) is likely to be violated, since $\Delta y_{i2} = \eta_i + v_{i2}$ in this case. One exception is the model considered by Arellano and Bover (1995), in which η_i is specified as $(1-\alpha)\eta_i^*$. In this case Δy_{it} is a pure random walk for each individual when $\alpha = 1$. This has two implications: first, $E(\Delta y_{it} y_{i,t-s}) = 0$ for $s \geq 1$, so that lagged instruments do not identify α in the standard first differenced equations; secondly, estimating α is straightforward in this case, since y_{i2} is uncorrelated with u_{i3} and so OLS is a consistent estimator in the levels equations for periods $2, ..., T$.[7].

In the more general model with $\alpha = 1$, Δy_{it} is a random walk with individual-specific drifts. In this case we notice that although conditions (4.2) and (4.1) are not valid, there are alternative valid linear moment conditions for the levels equations, namely

$$E(u_{it}(\Delta y_{i,t-1} - \Delta y_{i,t-2})) = 0; \text{ for } t = 4, 5, ..., T, \tag{4.5}$$

since $\Delta y_{i,t-1} - \Delta y_{i,t-2} = \Delta u_{i,t-1}$. When $\alpha = 1$ conditions (4.5) therefore coincide with conditions (2.10), which can be implemented in a linear GMM estimator for $T \geq 4$ without any further restrictions on the initial conditions in this case.

[7]A GMM estimator using instruments $y_{i,t-s}$ for $s \geq 1$ in the levels equations for periods $2, ..., T$ would be more efficient in the presence of heteroskedasticity.

4.2. Optimal GMM

The validity of the additional linear moment restrictions like (4.1) and (4.2) can often be regarded as an empirical issue, since the extra moment conditions are overidentifying restrictions that can be tested using standard procedures. In the Monte-Carlo results below we focus on the $\alpha < 1$ case. We evaluate the efficiency gains from correctly imposing restriction (4.2), the bias that results from incorrectly assuming (4.2), and the ability to detect such misspecification.

Calculation of the GMM estimators based on conditions (4.1) and (4.2) requires a stacked system comprising (T-2) equations in first differences and the (T-2) equations in levels corresponding to periods 3,...,T, for which instruments are observed. The instrument matrix for this system can be written

$$\mathbf{Z}_i^+ = \begin{bmatrix} \mathbf{Z}_i & 0 & 0 & \ldots & 0 \\ 0 & \Delta y_{i2} & 0 & \ldots & 0 \\ 0 & 0 & \Delta y_{i3} & \ldots & 0 \\ . & . & . & \ldots & 0 \\ 0 & 0 & 0 & \ldots & \Delta y_{i,T-1} \end{bmatrix} \tag{4.6}$$

where $\mathbf{Z}_i$ is as defined in section 2.[8] The calculation of the two-step GMM estimator is then analogous to that described above, and is detailed in Appendix B.

This linear GMM estimator based on the system of both differenced and levels equations exploits the additional restriction in (4.2), and is therefore asymptotically strictly more efficient than the non-linear GMM estimator exploiting (2.10) if this restriction is valid.

[8]The use of conditions (4.5) would produce a similar estimator, with levels equations for periods 4,...,T only and with $\Delta y_{i,t-1}$ replaced by $\Delta y_{i,t-1} - \Delta y_{i,t-2}$.

5. Monte Carlo Results

In this section we report the results of a Monte Carlo study which investigates the finite sample behaviour of the standard first differenced GMM estimator, the conditional GLS estimator and the GMM estimator that exploits the additional restriction (4.2) on the initial condition process. We document the improvement in precision and finite sample performance that may result from the use of these alternative estimators. This is found to be particularly important when the autoregressive parameter α is high and when the number of time series observations is small.

5.1. The Monte Carlo Design

In all experiments the series y_{it} was generated as

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

for $i = 1, 2, ..., N$ and $t = 2, 3, ..., T$, where $\eta_i \sim i.i.d.N(0, 1)$ and $v_{it} \sim i.i.d.N(0, 1)$. The initial conditions y_{i1} were generated from a model of the form

$$y_{i1} = k + \delta\eta_i + u_{i1}$$

where $u_{i1} \sim i.i.d.N(0, \sigma^2_{u1})$. New values for the initial observations are drawn in each of the replications.

5.2. Bias and Efficiency

Table 1(a) reports sample means and standard deviations from 1000 replications for a stationary design with $\alpha = 0.5$. The sample size used here has $N = 100$ and $T = 6$. The parameters for the initial conditions process are $k = 0$,

$\delta = 2$ and $\sigma_{u1}^2 = 4/3$. The variance of the initial conditions is chosen such that $var(y_{i1}) = var(y_{it})$ for $t = 2, ..., T$, so that the model is stationary.

The upper part of Table 1(a) reports results for the standard panel data estimators, which are presented for comparison. The first two rows report OLS in levels and Within Groups. As expected, OLS in levels results in a serious upwards bias in the estimate of α due to the presence of permanent effects, whilst Within Groups results in a serious downwards bias due to the short time series.

The middle two rows report infeasible (Inf.) and feasible (F.) versions of the standard error components GLS estimator. Estimates of the variance components σ_η^2 and σ_v^2 are obtained using the residuals from the asymptotically most efficient one step GMM estimator that we calculate for each model - in this case this is GMM1 (SYS) as defined below. These standard error components GLS estimators result in a serious bias as a consequence of the correlation between y_{i1} and the permanent effects.

The last two rows report one step (GMM1 (DIF)) and two step (GMM2 (DIF)) Generalised Method of Moments estimators based on the T-2 first differenced equations and the moment conditions given in (2.7) only. These are the conditions which depend on the absence of serial correlation in the v_{it}. This is the standard differenced GMM estimator that has been widely adopted in dynamic panel data applications, and in contrast to the previous estimators these are consistent for this model. Both one step and two step versions show a downward finite sample bias. There is no gain to be expected from using the two step GMM estimator here, as there is no heteroskedasticity in the design, and no improvement is found.

The lower part of Table 1(a) reports results for the estimators presented in sections 3 and 4, which exploit information on the initial conditions. The first two rows report infeasible (Inf.) and feasible (F.) versions of the conditional GLS estimator described in section 3. The feasible CGLS estimator again uses residuals based on GMM1 (SYS) to estimate the required variance components.

The infeasible CGLS estimator shows no discernible bias and sets the benchmark for precision in this experiment. The infeasible CGLS estimator performs less well than the feasible version. This estimator has an upward finite sample bias, but offers a considerable improvement in precision relative to the differenced GMM estimators.[9]

The final two rows report results for GMM estimators based on the stacked system of T-2 first differenced equations and T-2 levels equations for $t = 3, 4, ..., T$, as described in section 4. These estimators use the moment conditions in (2.7) and those in (4.1) and (4.2). Recall that the validity of these instruments for the levels equations depends on the restriction on the initial conditions process given in (4.2), which is satisfied in this stationary design. GMM1 (SYS) is a one step GMM estimator for the system which uses the matrix $\tilde{\mathbf{H}}$ (see Appendix B), with $\gamma = 0$, in the construction of the weight matrix $\mathbf{A}_N$. GMM2 (SYS) is a standard two step estimator for the stacked system, using the residuals from GMM1 in constructing the optimal weight matrix.

These estimators perform considerably better than those based on the first differenced equations alone. The information in the levels equations helps to reduce the finite sample bias a little, and to improve the precision considerably. The two step estimator has a smaller bias than the one step estimator, but a slightly

[9]The performance of the feasible CGLS estimator may partly reflect the imprecision of the one step GMM estimator in this design (recall that the correct GLS transformation is needed for the consistency of this estimator), and it may be that iterating on the feasible conditional GLS estimator would result in better finite sample performance.

bigger standard deviation - even though the two step estimator is asymptotically efficient relative to the one step estimator in this system. Both GMM estimators perform marginally better than the feasible CGLS estimator, which does not exploit restriction (4.2).

One further point to report concerns the behaviour of the asymptotic standard errors associated with these GMM estimators. Arellano and Bond (1991) noted a serious downward bias in the asymptotic standard errors for the two step GMM estimator based on the first-differenced equations. In Table 1(a), the mean of the asymptotic standard errors for GMM2 (DIF) was 0.1095, only 82% of the empirical standard deviation. This small sample bias increases with the number of overidentifying restrictions. For GMM2 (SYS) the mean of the asymptotic standard errors was only 66% of the corresponding empirical standard deviation. This problem is much less serious for the one-step GMM estimators, which may therefore allow more reliable inference in small samples[10].

5.3. Relaxing Stationarity

Table 1(b) reports means and standard deviations for the same set of estimators from an experiment where stationarity is not imposed. The parameters of the initial conditions process in this case are $k = 0$, $\delta = 0$, and $\sigma_{u1}^2 = 16/3$, so that the initial conditions are simply noise, with $y_{i1} \sim i.i.d.N(0, 16/3)$. In this case there is no correlation between y_{i1} and η_i, so that the standard error components GLS estimator is consistent here. Moreover this implies that there is correlation between Δy_{i2} and η_i, so that condition (4.2) is violated and there are no linear

[10]The mean of the asymptotic standard errors was respectively 98.2% and 99.7% of the empirical standard deviation for GMM1 (DIF) and GMM1 (SYS).

moment conditions available in the levels equations. As in Table 1(a), the results are based on 1000 replications for a sample size $N = 100$ and $T = 6$, and the true value of α is again 0.5.

Here the standard GLS estimator performs very well, although it is perhaps more interesting that the conditional GLS estimator performs only marginally less well. The much better relative performance of the feasible GLS estimators, compared to Table 1(a), is probably due to the improvement in the performance of the one step GMM estimator used to estimate the variance components that results from this change in the initial conditions process[11]. The GMM estimator based on the first differenced equations appears satisfactory here, but has both a larger bias and a larger standard deviation than the feasible GLS estimators. The GMM(SYS) estimators based on the stacked system of equations are seriously biased, as would be expected for this specification of the initial conditions.

It is useful to note that standard GMM specification tests to appear to have power to detect the invalidity of these additional linear moment conditions in this non-stationary design. The conventional Sargan/Hansen test of over-identifying restrictions rejects the validity of the GMM2 (SYS) estimator in 992 cases at the 90% level; 971 cases at the 95% level; and 810 cases at the 99% level. The corresponding rejection frequencies for the consistent GMM2 (DIF) estimator are 110, 45 and 6 respectively.

5.4. Relaxing Normality and Homoskedasticity

Table 2 considers the performance of these estimators in situations where the disturbances are not homoskedastic normal. In Table 2(a) the errors v_{it} are

[11]The estimates of these variance components are obtained from GMM1 (DIF) in this case.

drawn from a $\chi^2(1)$ distribution, centred to have mean zero and scaled to have variance one. Hence the errors are highly skewed. Otherwise the design is the same as Table 1(a). A comparison of Tables 1(a) and 2(a) suggests that this non-normality has very little effect on the behaviour of these estimators, although the relative performance of the two step GMM estimators is marginally better in this case.

In Table 2(b) the errors are normal but heteroskedastic, with the variance of v_{it} proportional to $y^2_{i,t-1}$ and scaled to have variance one on average across the sample. Otherwise the design is again similar to that in Table 1(a). The introduction of this heteroskedasticity has a much bigger impact on the results. The precision of all the estimators considered here deteriorates. There is a significant increase in the small sample bias found for the standard GMM estimators based only on first differenced equations. This effect is much smaller for the GMM(SYS) estimators based on stacked levels and differenced equations, and also for the feasible CGLS estimator. As might be expected, the relative performance of the two step GMM estimators also improves when heteroskedasticity is present, and GMM2(SYS) outperforms feasible CGLS in terms of both bias and precision in this experiment.

Both non-normality and heteroskedasticity also resulted in a considerable increase in the downward bias of the asymptotic standard errors for the two step GMM estimators. For GMM2 (SYS), the mean of the asymptotic standard errors was respectively 55% and 42% of the corresponding empirical standard deviations in Tables 2(a) and 2(b). Asymptotic inference based on these estimators could be seriously misleading when non-normality or heteroskedasticity is suspected.

Again this problem was much less serious for the one step GMM estimators[12].

5.5. Efficiency Gains

Table 3 summarises the gains in precision in a range of experiments when the two step GMM estimator based on the stacked system of levels and differenced equations is compared to the standard two step GMM estimator based on differenced equations alone. This is shown as the ratio of the empirical variance of the differenced estimator in our simulation experiments to the empirical variance of the system estimator. This is reported for a range of sample sizes and values of the autoregressive parameter α. In each case the design of the initial conditions process is chosen to satisfy stationarity, so the linear moment conditions (4.1) and (4.2) are valid for the equations in levels.

Table 3(a) reports results for the case of $T = 4$. For high values of the autoregressive parameter α we find dramatic efficiency gains compared to the basic first-differenced GMM estimator. Lagged levels of the series provide weak instruments for first differences in these cases. Not surprisingly, therefore, the gain in precision that results from exploiting the initial condition information decreases for lower values of α.

Table 3(b) reports results for the case of $T = 11$. Again we find big efficiency improvements for high values of α. However, comparing Tables 3(a) and 3(b), we find that for a given value of α, the gain from exploiting the initial conditions tends to decrease as the number of time series observations gets larger.

We can compare these simulation results with ratios of the corresponding (large

[12]For example, the mean of the asymptotic standard errors for GMM1 (SYS) was 99% and 93% of the empirical standard deviations in Tables 2(a) and 2(b).

N) asymptotic variances. For $T = 4$ and $\alpha = 0.8$ the asymptotic variance ratio is 13.45, and for $\alpha = 0.9$ this is 49.51, which are close to the Monte Carlo results found in Table 3(a) for $N = 500$. These asymptotic results for our GMM estimator can also be compared to results reported by Ahn and Schmidt for the GMM estimator which exploits the non-linear moment conditions (2.10) but not the linear conditions (4.1) and (4.2) discussed in section 4. For $T = 4$ and with $\alpha = 0.8$ and $\alpha = 0.9$ Ahn and Schmidt (1995) report asymptotic variance ratios of 2.42 and 2.54 respectively. This comparison indicates that exploiting the extra initial conditions information (4.2) can, when valid, result in a dramatic improvement in precision at low values of T.

In Table 4 we further investigate the finite sample performance of these GMM estimators. For low values of T and high values of α we find both a huge downward bias and very low precision for the standard first-differenced estimator, even when $N = 500$. The initial condition information exploited in our system estimator not only greatly improves the precision but also greatly reduces the finite sample bias in this case. We can also observe that whilst the finite sample bias of the differenced estimator is generally downwards (in the direction of OLS in first differences), the finite sample bias of the system estimator is generally upwards (in the direction of OLS levels). However the advantages of exploiting the additional moment conditions when α is high are manifest in the results of these simulations.

6. Summary and Conclusions

In this paper we have discussed the importance of exploiting initial condition information in generating efficient estimators for dynamic panel data models where the number of time-series observations is small. We have focused on the individual effects autoregressive model $y_{it} = \alpha y_{i,t-1} + \eta_i + v_{it}$ although our results easily extend to dynamic models with regressors.

We considered two estimators that can improve the precision of the standard first-differenced GMM estimator for this model. One approach conditions on the observed initial values to obtain a system that can be estimated consistently by error-components GLS. This estimator was shown to be equivalent to the optimal non-linear GMM estimator in the normal, homoskedastic model. The second approach imposes an additional restriction on the initial conditions process, under which all the moment conditions available can be exploited by a linear GMM estimator in a system of first-differenced and levels equations. This estimator is strictly more efficient than non-linear GMM when the additional restriction is valid.

The finite sample properties of these estimators were studied using Monte Carlo simulations. Both can improve dramatically on the performance of the usual first-differenced GMM estimator when the autoregressive parameter is moderately high and the number of time-series observations is moderately small. In addition, asymptotic variance comparisons suggest that our system GMM estimator can be considerably more efficient than non-linear GMM in this case.

In models with regressors x_{it}, the precise set of moment conditions available will depend on assumptions made about the correlation between x_{it} and u_{it}. To exploit the linear moment conditions in equations (4.1) and (4.2) an additional requirement is that $E(\Delta x_{it}\eta_i) = 0$, which is clearly weaker than requiring the levels of x_{it} to be uncorrelated with the individual effects. This is precisely the case considered in Arellano and Bover (1995). Together with the restriction (4.2) on the initial conditions process this allows the use of both lagged Δy_{it} and lagged Δx_{it} as instruments in levels equations. Stricter exogeneity restrictions on the x_{it} will allow the use of further instruments, as detailed in Arellano and Bond (1991), Ahn and Schmidt (1995) and Arellano and Bover (1995).

The extension of the conditional GLS estimator to models with regressors requires that the x_{it} are strictly exogenous with respect to u_{it}. Rather than conditioning on y_{i1} alone, in this case the residuals from the linear projection of y_{i1} on the vector $(x_{i1}, x_{i2}, ..., x_{iT})$ should be included in the conditional model (3.4). Details of the resulting estimator and some extensions to models with autocorrelation and heteroskedasticity over time can be found in Blundell and Smith (1991).

Appendix A: Proof of Proposition 3.1.

The conditional independence assumption underlying the normal linear model allows us to write the optimal moment restrictions as

$$E(\mathbf{u}(\mathbf{y},\alpha)|\mathbf{Z}) = \mathbf{0}.$$

If we define

$$\mathbf{A} = E(\mathbf{Z}'\mathbf{u}(\mathbf{y},\alpha)\mathbf{u}(\mathbf{y},\alpha)'\mathbf{Z}|\mathbf{Z}) \text{ and } \mathbf{G} = E(\mathbf{Z}'\partial\mathbf{u}(\mathbf{y},\alpha)/\partial\alpha|\mathbf{Z})$$

then the optimal GMM estimator uses the instruments $\mathbf{G}'\mathbf{A}^{-1}$, see McFadden and Newey (1994).

Note that

$$E(\mathbf{Z}'\mathbf{u}(\mathbf{y},\boldsymbol{\theta})|\mathbf{z}) = \int \mathbf{Z}'\mathbf{u}(\mathbf{y},\boldsymbol{\theta})\mathbf{f}(\mathbf{y}|\mathbf{z},\boldsymbol{\theta})\mathbf{dy} = \mathbf{0},$$

integrating by parts we find

$$E(\mathbf{Z}'\partial\mathbf{u}(\mathbf{y},\boldsymbol{\theta})/\partial\boldsymbol{\theta}|\mathbf{z}) = -\, E(\mathbf{Z}'\mathbf{u}(\mathbf{y},\boldsymbol{\theta})\partial\ln\mathbf{f}(\mathbf{y}|\mathbf{z},\boldsymbol{\theta})/\partial\boldsymbol{\theta}'|\mathbf{z})).$$

It follows that

$$\mathbf{G} = \; - \, E(\mathbf{Z}'\mathbf{u}(\mathbf{y},\alpha)\mathbf{s}(\mathbf{y},\alpha)'|\mathbf{Z})$$

where $\mathbf{s}(\mathbf{y},\alpha)$ is the score of the *conditional* loglikelihood. Since this score can be written

$$\mathbf{s}(\mathbf{y},\alpha) = -\mathbf{C}\mathbf{Z}'\mathbf{u}(\mathbf{y},\alpha)$$

we have $\mathbf{G} = \mathbf{A}\mathbf{C}'$ or $\mathbf{C} = \mathbf{G}'\mathbf{A}^{-1}$. Thus under normality the CGLS estimator is asymptotically equivalent to the optimally weighted non-linear GMM.

Appendix B: The choice of weight matrix in the initial conditions GMM estimator.

Corresponding to the use of the weight matrix $\mathbf{H}$ in calculating the one-step GMM estimator in the usual first-differenced case (see footnote 1), for the system of first-differenced and levels equations we consider the weight matrix

$$\bar{\mathbf{H}} = \begin{bmatrix} \mathbf{H} & \mathbf{L} \\ \mathbf{L}' & \mathbf{\Gamma} \end{bmatrix} \tag{6.1}$$

where

$$\mathbf{\Gamma} = \begin{bmatrix} 1+\gamma & \gamma & \gamma & \dots & \gamma \\ \gamma & 1+\gamma & \gamma & \dots & \gamma \\ \gamma & \gamma & 1+\gamma & \dots & \gamma \\ . & . & . & \dots & . \\ \gamma & \gamma & \gamma & \dots & 1+\gamma \end{bmatrix}, \tag{6.2}$$

$\gamma = \sigma_\eta^2/\sigma_\nu^2$ and

$$\mathbf{L} = \begin{bmatrix} 1 & 0 & 0 & \dots & 0 \\ -1 & 1 & 0 & \dots & 0 \\ 0 & -1 & 1 & \dots & 0 \\ . & . & . & \dots & . \\ 0 & 0 & 0 & \dots & 1 \end{bmatrix}. \tag{6.3}$$

Notice that $\bar{\mathbf{H}}$ depends on the unknown variance components parameters σ_η^2 and σ_ν^2 and so cannot be implemented in one step. A one-step estimator can be obtained by setting $\gamma = 0$, but in contrast to the usual first-differenced case this one-step estimator will not be asymptotically equivalent to the two-step estimator even in the i.i.d. case.

References

[1] Ahn, S.C. and P.Schmidt (1995), "Efficient estimation of models for dynamic panel data", *Journal of Econometrics*, 68, 5-28.

[2] Anderson, T.W. and C. Hsaio (1981),"Estimation of Dynamic Models with Error Components", *Journal of the American Statistical Association*, 76, 598-606.

[3] Arellano, M. and S.R.Bond (1991), "Some tests of specification for panel data: Monte Carlo evidence and an application to employment equations", *Review of Economic Studies*, 58, 277-297.

[4] Arellano, M. and O.Bover (1995), "Another look at the instrumental-variable estimation of error-components models", *Journal of Econometrics*, 68, 29-52.

[5] Bhagarva, A. and J.D. Sargan (1983),"Estimating Dynamic Random Effects Models from Panel data Covering Short Time Periods", *Econometrica*, 51, 1635-1659.

[6] Blundell, R.W. and R.J.Smith (1991), "Initial conditions and efficient estimation in dynamic panel data models", *Annales d'Economie at de Statistique*, 20/21, 109-123.

[7] Chamberlain, G. (1987), "Asymptotic efficiency in estimation with conditional moment restrictions", *Journal of Econometrics*, 34, 305-334.

[8] Crepon, B., Kramarz, F. and Trognon, A. (1993),"Parameter of Interest, Nuisance Parameter and Othogonality Conditions: An Application to Autoregressive Error Components Models ", CREST Working Paper, No. 9335, July.

[9] Hansen, L.P. (1982), "Large sample properties of generalised method of moment estimators", *Econometrica*, 50, 1029-1054.

[10] Holz-Eakin, D., W. Newey, and H.S.Rosen (1988),"Estimating Vector Autoregressions with Panel Data", *Econometrica*, 56, 1371-1396.

[11] Horowitz, J.L. (1995),"Bootstrap Methods in Econometrics: Theory and Numerical Performance" Working Paper #95-10, Department of Economics, Universitity of Iowa, July.

[12] Hsiao, C. (1986), *Analysis of Panel Data*, Cambridge: Cambridge University Press.

[13] McFadden, D. and Newey, W. K. and McFadden, D.(1995), "Large Sample Estimation and Hypothesis Testing", in Engle, R.F. and McFadden, D. (eds), *Handbook of Econometrics*, Vol IV, North-Holland.

[14] Nickell, S.J. (1981), "Biases in dynamic models with fixed effects", *Econometrica*, 49, 1417-1426.

[15] Prucha, I.R. (1984), "On the Asymptotic Efficiency of Feasible Aitken Estimators for Seemingly Unrelated Regression Models with Error Components", *Econometrica*, 52, 203-207.

[16] Sevestre, P. and A.Trognon (1985), "A note on autoregressive error components models", *Journal of Econometrics*, 29, 231-245.

Table 1: Varying the Initial Condition Processes

$\alpha = 0.5$, $N = 100$, $T = 6$

	(a)		(b)	
Estimator	**Mean**	**Std Dev**	**Mean**	**Std Dev**
OLS	0.8733	0.0209	0.7784	0.0348
Within	0.1693	0.0465	0.3519	0.0363
Inf. GLS	0.5842	0.0370	0.4993	0.0303
F. GLS	0.6458	0.0934	0.4967	0.0435
GMM1 (DIF)	0.4487	0.1219	0.4871	0.0574
GMM2 (DIF)	0.4523	0.1338	0.4869	0.0629
Inf. CGLS	0.5032	0.0424	0.4977	0.0314
F. CGLS	0.5664	0.0926	0.4971	0.0458
GMM1 (SYS)	0.5540	0.0886	0.7884	0.0812
GMM2 (SYS)	0.5342	0.0895	0.7655	0.0990

Replications = 1000

Initial Conditions:

(a) $y_{i1} = \frac{\eta_i}{1-\alpha} + u_{i1},\ u_{i1} \sim i.i.d.N(0, 4/3)$

(b) $y_{i1} \sim i.i.d.N(0, 16/3)$

Table 2: Introducing Non-Normalilty and Heteroskedasticity

$\alpha = 0.5$, $N = 100$, $T = 6$

$y_{i1} = \frac{\eta_i}{1-\alpha} + u_{i1}$

$u_{i1} \sim i.i.d.N(0, 4/3)$

	(a)		(b)	
Estimator	**Mean**	**Std Dev**	**Mean**	**Std Dev**
OLS	0.8715	0.0246	0.8699	0.0341
Within	0.1696	0.0567	0.1557	0.0741
Inf. GLS	0.5829	0.0447	0.5746	0.0587
F. GLS	0.6399	0.0885	0.6640	0.1322
GMM1 (DIF)	0.4456	0.1450	0.3181	0.1950
GMM2 (DIF)	0.4640	0.1411	0.3294	0.1817
Inf. CGLS	0.5046	0.0576	0.4952	0.0780
F. CGLS	0.5526	0.0886	0.5970	0.1436
GMM1 (SYS)	0.5543	0.0847	0.5717	0.1273
GMM2 (SYS)	0.5396	0.0811	0.5425	0.1116

Replications = 100

Error Distributions:

(a) $v_{it} = (\varepsilon_{it} - 1)/\sqrt{0.5}$ $where$ $\varepsilon_{it} \sim \chi^2(1)$

(b) $v_{it} \sim N(0, \sigma_{it}^2)$ $where$ $\sigma_{it}^2 = \left(\frac{y_{i,t-1}^2}{var(y)}\right)$

Table 3: Efficiency Gains

$$y_{i1} = \frac{\eta_i}{1-\alpha} + u_{i1}$$

$$u_{i1} \sim i.i.d.N(0, v_1)$$

v_1 is chosen such that $var(y_{i1}) = var(y_{it})\ for\ t = 2, ..., T$

		(a) T=4	(b) T=11
N	α	$\frac{var(GMM-DIF)}{var(GMM-SYS)}$	$\frac{var(GMM-DIF)}{var(GMM-SYS)}$
100	0.0	1.93	1.12
	0.3	2.52	1.28
	0.5	4.25	1.59
	0.8	43.27	6.99
	0.9	36.80	60.90
200	0.0	2.19	1.08
	0.3	3.31	1.18
	0.5	5.41	1.28
	0.8	34.02	2.69
	0.9	30.62	23.13
500	0.0	1.51	1.42
	0.3	1.91	1.75
	0.5	2.68	2.16
	0.8	12.21	3.36
	0.9	58.60	10.98

Replications = 100

Table 4: Two Step GMM Estimators underlying Table 3

α	**DIF** mean (std dev)	**SYS** mean (std dev)
T=4, N=100		
0.5	0.4489 (.2561)	0.5212 (.1242)
0.9	0.1919 (.8007)	0.9614 (.1320)
T=4, N=500		
0.5	0.4863 (.1042)	0.5026 (.0636)
0.9	0.5917 (.6897)	0.9092 (.0901)
T=11, N=100		
0.5	0.4712 (.0630)	0.5797 (.0499)
0.9	0.6530 (.1636)	0.9763 (.0210)
T=11, N=500		
0.5	0.4967 (.0272)	0.5038 (.0185)
0.9	0.8214 (.0729)	0.9303 (.0220)