Initial conditions and moment restrictions in dynamic panel data models

Richard Blundell[a],*, Stephen Bond[b]

[a] *Institute for Fiscal Studies and Department of Economics, University College London, London WC1E 6BT, UK*
[b] *Institute for Fiscal Studies and Nuffield College, Oxford OX1 1NF, UK*

Abstract

Estimation of the dynamic error components model is considered using two alternative linear estimators that are designed to improve the properties of the standard first-differenced GMM estimator. Both estimators require restrictions on the initial conditions process. Asymptotic efficiency comparisons and Monte Carlo simulations for the simple AR(1) model demonstrate the dramatic improvement in performance of the proposed estimators compared to the usual first-differenced GMM estimator, and compared to non-linear GMM. The importance of these results is illustrated in an application to the estimation of a labour demand model using company panel data. © 1998 Elsevier Science S.A. All rights reserved.

JEL classification: C23

Keywords: Dynamic panel data; Error components; Weak instruments; Initial conditions; GMM

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 widely used linear generalised method of moments (GMM) estimator obtained after first differencing has been found to have large finite sample bias and poor precision in simulation studies (see Alonso-Borrego and Arellano, 1996). Lagged

---

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

levels of the series provide weak instruments for first differences in this case. Here we consider two alternative estimators that impose further restrictions on the initial conditions process, designed to improve the properties of the standard first-differenced GMM estimator.

The first type of restriction justifies the use of an extended linear GMM estimator that 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 (see Arellano and Bover, 1995). Monte Carlo simulations and asymptotic variance calculations show that this extended GMM estimator offers dramatic efficiency gains in the situations where the basic first-differenced GMM estimator performs poorly. This estimator is also shown to encompass the GMM estimator based on the non-linear moment conditions available in the dynamic error components model (see Ahn and Schmidt, 1995), and we find substantial asymptotic efficiency gains relative to this non-linear GMM estimator.

The second type of restriction validates the use of the error components GLS estimator on an extended model that conditions on the observed initial values. This provides a consistent estimator under homoskedasticity which, under normality, is asymptotically equivalent to conditional maximum likelihood (see Blundell and Smith, 1991). A Monte Carlo analysis also suggests that this estimator has good finite sample properties. However, the conditional GLS estimator requires homoskedasticity, and only extends to a model with regressors if the regressors are strictly exogenous. This is not the case for the GMM estimators.

Both types of restrictions are satisfied under stationarity but both are also valid under weaker assumptions. The gain in precision that results from exploiting these initial condition restrictions in these two alternative estimators is shown to increase for higher values of the autoregressive parameter and as the number of time series observations gets smaller. Our Monte Carlo analysis finds both a large downward bias and very low precision for the standard first-differenced estimator in these cases. The initial condition information not only greatly improves the precision but also greatly reduces the finite sample bias.

The main contributions of the paper are: to characterise the weak instruments problem for the first-differenced GMM estimator in terms of the concentration parameter (cf. Staiger and Stock, 1997); to demonstrate that the levels restrictions suggested by Arellano and Bover (1995) remain informative in the cases where the first-differenced instruments become weak; to relate these restrictions explicitly to the initial conditions process; to evaluate the asymptotic efficiency gains that result from exploiting these restrictions compared to both the differenced GMM estimator and the non-linear GMM estimator of Ahn and Schmidt (1995); to relate the conditions needed for consistency of the Blundell and Smith (1991) conditional GLS estimator to the initial conditions process; and finally to evaluate the performance of these estimators using both Monte Carlo simulations and an application to company panel data.

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 (1981), Holtz-Eakin et al. (1988), Arellano and Bond (1991), and Ahn and Schmidt (1995). In Section 3 we evaluate the problem of weak instruments in the first-differenced instrumental variable estimator. Section 4 goes on to consider restrictions on the initial condition process that render lagged values of Δy_{it} valid as instruments for the levels equations and discusses the extended GMM estimator which is available when these restrictions are satisfied. In Section 5 we consider the conditional GLS estimator in which initial conditions are explicitly added to the model. Section 6 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 7 discusses the extensions to models with strictly exogenous and predetermined regressors. An application to a panel data model of labour demand is presented which illustrates the usefulness of the extended GMM estimator in practice. Section 8 concludes.

2. The model

We consider an autoregressive panel data model of the form

$$y_{it} = \alpha y_{i,t-1} + \beta'_1 x_{it} + \beta'_2 x_{it-1} + \eta_i + v_{it} \tag{2.1}$$

for $i = 1, \ldots, N$ and $t = 2, \ldots, T$, where $u_{it} \equiv \eta_i + v_{it}$ is the usual 'fixed effects' decomposition of the error term; N is large, T is fixed and $|\alpha| < 1$.[1] This has the corresponding 'common factor' restricted $(\beta_2 = -\alpha\beta_1)$ form

$$y_{it} = \beta'_1 x_{it} + f_i + \zeta_{it}$$

with

$$\zeta_{it} = \alpha\zeta_{i,t-1} + v_{it} \quad \text{and} \quad \eta_i = (1 - \alpha)f_i. \tag{2.2}$$

In our application to panel data employment equations presented in Section 7 we allow for the inclusion of x_{it} regressors, but for the evaluation of the various estimators we use an AR(1) model with unobserved individual-specific effects

$$y_{it} = \alpha y_{i,t-1} + \eta_i + v_{it} \tag{2.3}$$

----

[1] All of the estimators discussed below and their properties extend in an obvious fashion to higher-order autoregressive models.

for $i = 1, \ldots, N$ and $t = 2, \ldots, T$. Since our focus is on the role of initial conditions we will assume that η_i and v_{it} are independently distributed across i and have the familiar error components structure in which

$$\text{E}(\eta_i) = 0, \ \text{E}(v_{it}) = 0, \ \text{E}(v_{it}\eta_i) = 0 \quad \text{for } i = 1, \ldots, N \text{ and } t = 2, \ldots, T \qquad (2.4)$$

and

$$\text{E}(v_{it}v_{is}) = 0 \quad \text{for } i = 1, \ldots, N \text{ and } \forall t \neq s. \qquad (2.5)$$

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

$$\text{E}(y_{i1}v_{it}) = 0 \quad \text{for } i = 1, \ldots, N \text{ and } t = 2, \ldots, T. \qquad (2.6)$$

Conditions (2.4), (2.5) and (2.6) imply moment restrictions that are sufficient to (identify and) estimate α for $T \geqslant 3$.

3. Estimation in first differences

3.1. The standard moment conditions

In the absence of any further restrictions on the process generating the initial conditions, the autoregressive error components model (2.3)–(2.6) implies the following $m = 0.5(T - 1)(T - 2)$ orthogonality conditions which are linear in the α parameter

$$\text{E}(y_{i,t-s}\Delta v_{it}) = 0 \quad \text{for } t = 3, \ldots, T \text{ and } s \geqslant 2 \qquad (3.1)$$

where $\Delta v_{it} = v_{it} - v_{i,t-1}$. These depend only on the assumed absence of serial correlation in the time-varying disturbances v_{it}, together with the restriction in Eq. (2.6).

The moment restrictions in Eq. (3.1) can be expressed more compactly as

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

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

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

and $\bar{\mathbf{u}}_i$ is the $(T - 2)$ vector $(\Delta v_{i3}, \Delta v_{i4}, \ldots, \Delta v_{iT})'$.

The generalised method of moments estimator based on these moment conditions minimises the quadratic distance $(\bar{\mathbf{u}}'\mathbf{Z}A_N\mathbf{Z}'\bar{\mathbf{u}})$ for some metric A_N, where $\mathbf{Z}'$ is the $m \times N(T - 2)$ matrix $(\mathbf{Z}_1', \mathbf{Z}_2', \ldots, \mathbf{Z}_N')$ and $\bar{\mathbf{u}}'$ is the $N(T - 2)$

vector $(\bar{\boldsymbol{u}}_1', \bar{\boldsymbol{u}}_2', \ldots, \bar{\boldsymbol{u}}_N')$. This gives the GMM estimator for α as

$$\hat{\alpha}_{\text{dif}} = (\bar{\boldsymbol{y}}_{-1}' \boldsymbol{Z} \boldsymbol{A}_N \boldsymbol{Z}' \bar{\boldsymbol{y}}_{-1})^{-1} \bar{\boldsymbol{y}}_{-1}' \boldsymbol{Z} \boldsymbol{A}_N \boldsymbol{Z}' \bar{\boldsymbol{y}}. \tag{3.2}$$

where $\bar{\boldsymbol{y}}_i$ is the $(T-2)$ vector $(\Delta y_{i3}, \Delta y_{i4}, \ldots, \Delta y_{iT})$, $\bar{\boldsymbol{y}}_{i,-1}$ is the $(T-2)$ vector $(\Delta y_{i2}, \Delta y_{i3}, \ldots, \Delta y_{i,T-1})$, and $\bar{\boldsymbol{y}}$ and $\bar{\boldsymbol{y}}_{-1}$ are stacked across individuals in the same way as $\bar{\boldsymbol{u}}$.

Alternative choices for the weights $\boldsymbol{A}_N$ give rise to a set of GMM estimators based on the moment conditions in Eq. (3.1), all of which are consistent for large N and finite T, but which differ in their asymptotic efficiency.[2] In general the optimal weights are given by

$$\boldsymbol{A}_N = \left(N^{-1} \sum_{i=1}^{N} \boldsymbol{Z}_i' \hat{\bar{\boldsymbol{u}}}_i \hat{\bar{\boldsymbol{u}}}_i' \boldsymbol{Z}_i \right)^{-1},$$

where $\hat{\bar{\boldsymbol{u}}}_i$ are residuals from an initial consistent estimator. We refer to this as the *two-step* GMM estimator.[3] In the absence of any additional knowledge about the process for the initial conditions, this estimator is asymptotically efficient in the class of estimators based on the linear moment conditions [Eq. (3.1)] (see Hansen, 1982; Chamberlain, 1987).

3.1.1. Homoskedasticity

Ahn and Schmidt (1995) show that additional linear moment conditions are available if the v_{it} disturbances are homoskedastic through time, i.e. if

$$\mathrm{E}(v_{it}^2) = \sigma_i^2 \quad \text{for } t = 2, \ldots, T. \tag{3.3}$$

This implies $T-3$ orthogonality restrictions of the form

$$\mathrm{E}(y_{i,t-2} \Delta v_{i,t-1} - y_{i,t-1} \Delta v_{it}) = 0 \quad \text{for } t = 4, \ldots, T \tag{3.4}$$

---

[2] These estimators are all based on the normalisation (2.3). Alonso-Borrego and Arellano (1996) consider a symmetrically normalised instrumental variable estimator based on the normalisation invariance of the standard LIML estimator.

[3] As a choice of $\boldsymbol{A}_N$ to yield the initial consistent estimator, Arellano and Bond (1991) suggest

$$\boldsymbol{A}_N = \left(N^{-1} \sum_{i=1}^{N} \boldsymbol{Z}_i' \boldsymbol{H} \boldsymbol{Z}_i \right)^{-1},$$

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

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

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.

and allows a further $T - 3$ columns to be added to the instrument matrix Z_i.[4] Calculation of the one-step and two-step GMM estimators then proceeds exactly as described above.

3.2. The problem of weak instruments

The instruments used in the standard first-differenced GMM estimator become less informative in two important cases. First, as the value of α increases towards unity, and second, as the relative variance of the fixed effects η_i increases. To examine this further consider the case with $T = 3$. For $T = 3$, moment conditions (3.1) reduce to a single orthogonality condition, so that α is just-identified. The corresponding method of moments estimator $\hat{\alpha}_{\text{dif}}$ in Eq. (3.2) reduces to a simple instrumental variable (IV) estimator with reduced form (instrumental variable regression) equation

$$\Delta y_{i2} = \pi y_{i1} + r_i \quad \text{for } i = 1, \dots, N. \tag{3.5}$$

For sufficiently high α or variance of η_i, the least-squares estimate of the reduced form coefficient π can be made arbitrarily close to zero. In this case the instrument y_{i1} is only weakly correlated with Δy_{i2}. To see this notice that the model (2.3) implies that

$$\Delta y_{i2} = (\alpha - 1)y_{i1} + \eta_i + v_{i2} \quad \text{for } i = 1, \dots, N. \tag{3.6}$$

The least-squares estimator of $(\alpha - 1)$ in Eq. (3.6) is generally biased upwards, towards zero, since we expect $E(y_{i1}\eta_i) > 0$. Assuming stationarity and letting $\sigma_\eta^2 = \text{var}(\eta_i)$ and $\sigma_v^2 = \text{var}(v_{it})$, the plim of $\hat{\pi}$ is given by

$$\text{plim}\, \hat{\pi} = (\alpha - 1)\frac{k}{(\sigma_\eta^2/\sigma_v^2) + k} \quad \text{with } k = \frac{(1 - \alpha)^2}{(1 - \alpha^2)}.$$

The bias term effectively scales the estimated coefficient on the instrumental variable y_{i1} toward zero. We find that $\text{plim}\, \hat{\pi} \to 0$ as $\alpha \to 1$ or as $(\sigma_\eta^2/\sigma_v^2) \to \infty$, which are the cases in which the first stage F-statistic is $O_p(1)$. A graph showing both $\text{plim}\, \hat{\pi}$ and $\alpha - 1$ against α is given in Fig. 1, for $\sigma_\eta^2 = \sigma_v^2 = 1$, $T = 3$.

Here we have considered taking α towards unity with σ_η^2 fixed ($= \sigma_v^2$). This seems reasonable for the case where we are asking what happens to the properties of this IV estimator for specific values of α. An alternative is to

---

[4] The additional columns Z_{Hi} are

$$Z'_{Hi} = \begin{pmatrix} y_2 & -y_3 & 0 & \dots & 0 & 0 \\ 0 & y_3 & -y_4 & \dots & 0 & 0 \\ . & . & . & \dots & . & . \\ 0 & 0 & 0 & \dots & y_{T-2} & -y_{T-1} \end{pmatrix}.$$
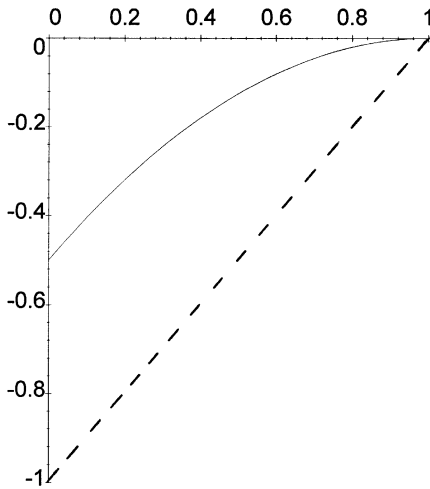
Fig. 1.   plim $\hat{\pi}$ and $\alpha - 1, \sigma_\eta^2$ fixed.

analyse the limiting sequence of $\alpha \to 1$ keeping $\sigma_\eta^2/(1-\alpha)^2$ fixed.[5] In this case the plim $\hat{\pi}$ has the form illustrated in Fig. 2, and although there is still substantial bias towards zero, it is less serious than in Fig. 1.[6]

We are interested in inferences using this differenced IV estimator when π is local to zero, that is where the instrument y_{i1} is only weakly correlated with Δy_{i2}. Following Nelson and Startz (1990a,b) and Staiger and Stock (1997) we characterise this problem of weak instruments using the concentration parameter. First note that the F-statistic for the first stage instrumental variable regression converges to a noncentral chi-squared with one degree of freedom. The concentration parameter is then the corresponding noncentrality parameter which we label τ in this case. The IV estimator performs poorly when τ approaches zero. Assuming stationarity, τ has the following simple characterisation in terms of the parameters of the AR model

$$\tau = \frac{(\sigma_v^2 k)^2}{\sigma_\eta^2 + \sigma_v^2 k} \quad \text{where } k = \frac{(1-\alpha)^2}{(1-\alpha^2)}. \tag{3.7}$$

The performance of the $\hat{\alpha}_{\text{dif}}$ estimator in this AR(1) specification can therefore be seen to deteriorate as $\alpha \to 1$, as well as for decreasing values of σ_v^2 and for increasing values of σ_η^2. To examine this further, Fig. 3 provides a plot of

---

[5] This would be equivalent to fixing the variance of f_i in the common factor specification of the model (2.2).

[6] Fig. 2 shows the case where $\sigma_\eta^2/(1-\alpha)^2 = \sigma_v^2 = 1$, $T = 3$.
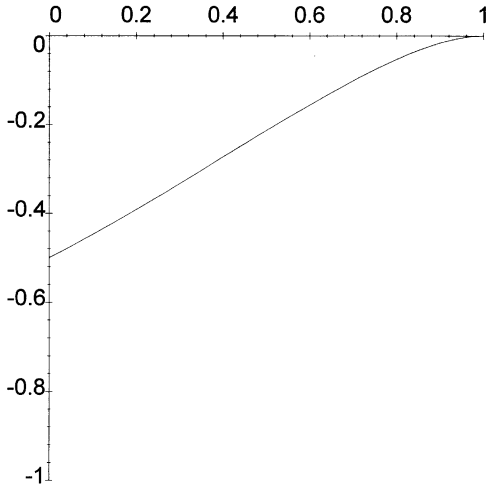
Fig. 2.   plim $\hat{\pi}, \sigma_\eta^2/(1-\alpha)^2$ fixed.
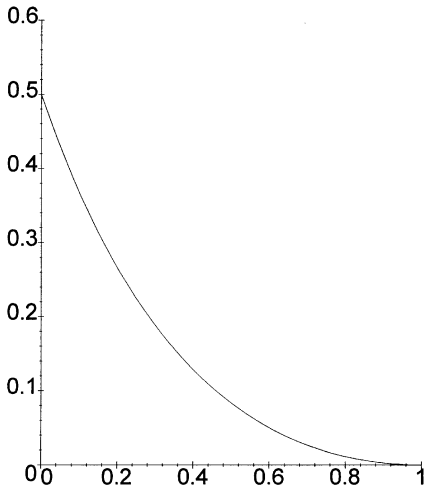
Fig. 3.   Concentration parameter τ for $T = 3$.

τ against α for the case $\sigma_\eta^2 = \sigma_v^2 = 1$, $T = 3$. We further investigate the finite sample properties of the differenced GMM estimator using Monte Carlo experiments in Section 6.

   In this special case, it is also interesting to note the following simple relationship between the asymptotic standard error of $\hat{\alpha}_{\text{dif}}$ and the concentration

parameter

$$\text{asyse}(\hat{\alpha}_{\text{dif}}) = \frac{\sigma_v}{\sqrt{2\tau\sigma_{\text{r}}^2}}$$

where σ_{r}^2 is the reduced form error variance $\text{E}(r_i^2)$ in Eq. (3.5). We see that the asymptotic variance of the $\hat{\alpha}_{\text{dif}}$ estimator explodes as $\tau \to 0$. For higher values of T the precision of $\hat{\alpha}_{\text{dif}}$ will also be shown to deteriorate as $\alpha \to 1$, although at a slower rate. A comparison of asymptotic variances is given in more detail in the following section.

4. Non-linear moment conditions and restrictions on the initial conditions 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 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 or when σ_η^2/σ_v^2 is high, since we have seen that lagged levels will be weak instruments in the differenced equations in these cases.

4.1. Non-linear moment conditions

A number of authors have previously suggested using the additional T-3 non-linear moment conditions (see, for example, Ahn and Schmidt, 1995, Eq. (4))

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

which are implied by Eqs. (2.4)–(2.6) and which could be expected to improve efficiency. These conditions relate directly to the absence of serial correlation in v_{it} and do not require homoskedasticity. Under the homoskedasticity through time restriction (3.3), Ahn and Schmidt (1995, Eq. (11b)) show the existence of an additional non-linear moment condition

$$\text{E}(\bar{u}_i\Delta u_{i3}) = 0 \quad \text{where } \bar{u}_i = \frac{1}{T-1}\sum_{t=2}^{T} u_{it}. \tag{4.2}$$

Asymptotic efficiency comparisons reported in Ahn and Schmidt (1995) confirm that these non-linear moments are particularly informative in cases where α is close to unity and/or where σ_η^2/σ_v^2 is high.

4.2. Initial conditions

In contrast to the non-linear moment restrictions (4.1), we consider the following T-3 linear moment conditions:[7]

$$E(u_{it}\Delta y_{i,t-1}) = 0 \quad \text{for } t = 4, 5, \ldots, T. \tag{4.3}$$

The use of lagged differences as possible instruments for equations in levels was proposed by Arellano and Bover (1995). Clearly, Eq. (4.3) does not imply Eq. (4.1). However, notice that since Δy_{i2} is observed, there is an additional restriction available, namely

$$E(u_{i3}\Delta y_{i2}) = 0. \tag{4.4}$$

Note also that the validity of this extra moment condition depends on a restriction on the initial conditions process generating y_{i1}. Condition (4.4) has two important implications. First, combining Eq. (4.4) with the model for periods $2, \ldots, T$ set out in Eqs. (2.3)–(2.6) implies the validity of the linear moment restrictions in Eq. (4.3).[8] Second, combining Eq. (4.3) and Eq. (4.4) implies the non-linear restrictions in Eq. (4.1), and renders these redundant for estimation. Thus, the complete set of second-order moment restrictions implied by Eqs. (2.3)–(2.6) and Eq. (4.4) can be implemented as a linear GMM estimator.

To examine the conditions under which Eq. (4.4) will hold, we write y_{i1} as

$$y_{i1} = \frac{\eta_i}{1-\alpha} + u_{i1}. \tag{4.5}$$

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. Noting that Eq. (4.4) is equivalent to $E[(\eta_i + v_{i3})(v_{i2} + (\alpha - 1)u_{i1})] = 0$, necessary conditions for Eq. (4.4) are then

$$E(u_{i1}\eta_i) = E(u_{i1}v_{i3}) = 0 \quad \text{for } i = 1, \ldots, N. \tag{4.6}$$

The key requirement is therefore that the *deviations* of the initial conditions from $\eta_i/(1-\alpha)$ are uncorrelated 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 $\sum_{s=-1}^{\infty}(\alpha^{s+1}v_{i,-s})$, and where assumptions (2.4) and (2.5) are maintained for all s and t. However, stationarity is not necessary for the

---

[7] In this section we focus on moment conditions that remain valid under heteroskedasticity.

[8] Given Eqs. (2.3)–(2.6), we can write

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

and the $\Delta v_{i,t-s}$ are uncorrelated with η_i. Hence $E(\Delta y_{it}\eta_i) = \alpha^{t-2}E(\Delta y_{i2}\eta_i)$.
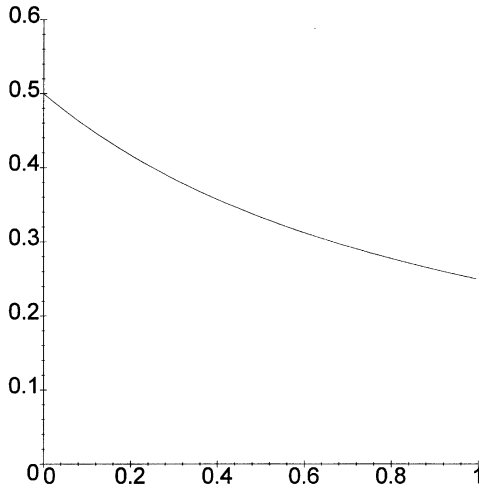
Fig. 4.   plim $\hat{\pi}$.

validity of the extra linear moment conditions (4.3) and (4.4). Condition (4.6) imposes no restriction on the variance of u_{i1}, and any entry period 'disequilibrium' from $\eta_i/(1 - \alpha)$ which is randomly distributed across agents will preserve condition (4.4). Other departures from stationarity such as $y_{i1} = k \forall i$ or $y_{i1} \sim$ i.i.d.$(0, \sigma_{y_1}^2)$ will violate Eq. (4.6) however, so this restriction is not without content.

To show that the moment conditions (4.3) and (4.4) remain informative when α approaches unity or σ_η^2/σ_v^2 becomes large, we again consider the case of $T = 3$. Here we can use one equation in levels

$$y_{i3} = \alpha y_{i2} + (\eta_i + v_{i3})$$

for which the instrument available is Δy_{i2}, and the reduced form equation is

$$y_{i2} = \pi \Delta y_{i2} + r_i.$$

In this case the plim $\hat{\pi}$ is given by

$$\text{plim } \hat{\pi} = \frac{1}{2}\left(\frac{1 - \alpha}{1 - \alpha^2}\right)$$

which is illustrated in Fig. 4.

As was the case for $\hat{\alpha}_{\text{dif}}$, the performance of the GMM estimator based on this levels equation can be seen to deteriorate as $\alpha \to 1$. However, with $|\alpha| < 1$, this moment condition stays informative in clear contrast to the $\hat{\alpha}_{\text{dif}}$ estimator. The efficient GMM estimator will combine both sets of orthogonality restrictions, to which we turn next.

4.3. A system GMM estimator

Calculation of the GMM estimators using conditions (4.3) and (4.4) can be based on a stacked system comprising all $(T-2)$ equations in first differences and the $(T-2)$ equations in levels corresponding to periods $3, \dots, T$, for which instruments are observed. The instrument matrix for this system can be written

$$
\mathbf{Z}_i^+ = \begin{bmatrix}
\mathbf{Z}_i & 0 & 0 & \cdots & 0 \\
0 & \Delta y_{i2} & 0 & \cdots & 0 \\
0 & 0 & \Delta y_{i3} & \cdots & 0 \\
\cdot & \cdot & \cdot & \cdots & 0 \\
0 & 0 & 0 & \cdots & \Delta y_{i,T-1}
\end{bmatrix} \tag{4.7}
$$

where $\mathbf{Z}_i$ is as defined in Section 3. The calculation of the two-step GMM estimator is then analogous to that described above.[9]

4.4. Asymptotic variance comparisons

To quantify the gains in asymptotic efficiency that result from exploiting the linear moment conditions (4.3) and (4.4), Table 1 reports the ratio of the asymptotic variance of the standard first-differenced GMM estimator described in Section 3.1 to the asymptotic variance of the system GMM estimator described in Section 4.3. These asymptotic variance ratios are calculated assuming both stationarity and homoskedasticity. They are presented for $T = 3$ and $T = 4$, for two fixed values of σ_η^2/σ_v^2, and for a range of values of the autoregressive parameter α. For comparison, we also reproduce from Ahn and Schmidt (1995) the corresponding asymptotic variance ratios comparing first-differenced GMM to the non-linear GMM estimator which uses the quadratic moment conditions (4.1), but *not* the extra linear moment conditions (4.3) and (4.4). In the $T = 3$ case there are no quadratic moment restrictions available. These calculations suggest that exploiting conditions (4.3) and (4.4) can result in dramatic efficiency gains when $T = 3$, particularly at high values of α and high values of σ_η^2/σ_v^2. These are indeed the cases where we find the instruments used to obtain the first-differenced estimator to be weak.

In the $T = 4$ case we still find dramatic efficiency gains at high values of α. Comparison to the results for the non-linear GMM estimator also shows that the gains from exploiting conditions (4.3) and (4.4) can be much larger than the gains from simply exploiting the non-linear restrictions (4.1).

---

[9] In this case, though, there is no one-step GMM estimator that is asymptotically equivalent to the two-step estimator, even in the special case of i.i.d. disturbances.

Table 1
Asymptotic variance ratios

	System	Non-linear	System	Non-linear
	$\sigma_\eta^2/\sigma_v^2 = 1$	$\sigma_\eta^2/\sigma_v^2 = 1$	$\sigma_\eta^2/\sigma_v^2 = 0.25$	$\sigma_\eta^2/\sigma_v^2 = 0.25$
α	$T = 3$		$T = 3$	
0.0	1.33	n/a	1.33	n/a
0.3	2.15		1.89	
0.5	4.00		2.91	
0.8	28.00		13.10	
0.9	121.33		47.91	
α	$T = 4$		$T = 4$	
0.0	1.75	1.67	1.40	1.29
0.3	2.31	1.91	1.77	1.33
0.5	3.26	2.10	2.42	1.35
0.8	13.97	2.42	8.88	1.41
0.9	55.40	2.54	30.90	1.45

In the Monte Carlo simulations presented in Section 6 we investigate whether similar improvements are found in finite samples.

4.5. Homoskedasticity

In the case where the initial conditions satisfy restriction (4.4) and the v_{it} satisfy restriction (3.3), Ahn and Schmidt (1995), (Eq. (12b)) show that the $T - 2$ homoskedasticity restrictions (3.4) and (4.2) can be replaced by a set of $T - 2$ moment conditions

$$E(y_{it}u_{it} - y_{i,t-1}u_{i,t-1}) = 0 \quad \text{for } t = 3, \dots, T \tag{4.8}$$

which are all linear in the parameter α. The non-linear condition (4.2) is again redundant for estimation given Eq. (4.4), and the complete set of second-order moment restrictions implied by Eqs. (2.3)–(2.6), (3.3) and (4.4) can be implemented in a linear GMM estimator.

5. The conditional GLS estimator

In the autoregressive error components model with homoskedasticity across *both* individuals and time,[10] restrictions on the initial conditions process can be used to derive a consistent *conditional* GLS (CGLS) estimator by including y_{i1} in

---

[10] In this section we require $E(v_{it}^2) = \sigma_v^2$ and $E(\eta_i^2) = \sigma_\eta^2$ for $i = 1, \dots, N$ and $t = 2, \dots, T$.

each of the $T - 1$ levels equations and then applying the standard error components GLS estimator. Under joint normality, this is the conditional ML estimator (CMLE) proposed by Blundell and Smith (1991).[11]

To examine this estimator we write the model for each i as a $T - 1$ equation system

$$\boldsymbol{y}_i = \alpha \boldsymbol{y}_{i,-1} + \boldsymbol{u}_i = \alpha \boldsymbol{y}_{i,-1} + \boldsymbol{l}\eta_i + \boldsymbol{v}_i \tag{5.1}$$

where $\boldsymbol{y}'_i = (y_{i2}, \ldots, y_{iT})$, $\boldsymbol{y}'_{i,-1} = (y_{i1}, \ldots, y_{i,T-1})$, $\boldsymbol{l}$ is the $T - 1$-unit vector, $\boldsymbol{u}'_i = (u_{i2}, \ldots, u_{iT})$ and $\boldsymbol{v}'_i = (v_{i2}, \ldots, v_{iT})$. The error component assumptions imply

$$\mathrm{E}(\boldsymbol{u}_i \boldsymbol{u}'_i) = \boldsymbol{\Omega} = \sigma^2_\eta \boldsymbol{J}_{T-1} + \sigma^2_v \boldsymbol{I}_{T-1}. \tag{5.2}$$

where $\boldsymbol{J}_{T-1}$ is the $(T-1) \times (T-1)$ unit matrix $\boldsymbol{ll}'$ and $\boldsymbol{I}_{T-1}$ the $(T-1) \times (T-1)$ identity matrix.

The presence of the fixed effect η_i in u_{it} not only implies

$$\mathrm{E}(\boldsymbol{y}'_{i,-1}\boldsymbol{u}_i) \neq \boldsymbol{0} \tag{5.3}$$

but also

$$\mathrm{E}(\boldsymbol{y}'_{i,-1}\boldsymbol{\Omega}^{-1}\boldsymbol{u}_i) \neq \boldsymbol{0}. \tag{5.4}$$

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).

Notice, however, that the system may be written

$$\boldsymbol{A}\boldsymbol{y}_i = \boldsymbol{a}y_{i1} + \boldsymbol{l}\eta_i + \boldsymbol{v}_i \tag{5.5}$$

with $\boldsymbol{A}$ a $(T-1) \times (T-1)$ lower triangular matrix and $\boldsymbol{a}' = (\alpha, 0, \ldots, 0)$. Since $\boldsymbol{A}$ is lower triangular the inconsistency of the GLS estimator can be seen to depend on $\mathrm{E}(y_{i1}\eta_i) \neq 0$. That is, although diagonalising the $(T-1) \times (T-1)$ error covariance matrix, premultiplying the $T - 1$ system (5.5) by $\boldsymbol{\Omega}^{-1/2}$ introduces y_{i1} into all $T - 1$ equations of the transformed system and induces a correlation with η_i. The aim of the *conditional* GLS estimator is to eliminate this correlation.

Given Eq. (2.6), we have that $\mathrm{E}(y_{i1}u_{it}) = \mathrm{E}(y_{i1}\eta_i)$ for $t = 2, \ldots, T$. We then make the following assumption

$$\mathrm{E}(\eta_i | y_{i1}) = \phi y_{i1} \quad \text{with } \mathrm{E}(y^2_{i1}) = \sigma^2_1 \quad \text{for } i = 1, \ldots, N. \tag{5.6}$$

Including y_{i1} in each equation of the system (5.5) gives an *extended* system

$$\boldsymbol{A}\boldsymbol{y}_i = \boldsymbol{a}y_{i1} + \boldsymbol{l}\phi y_{i1} + \boldsymbol{l}\tilde{\eta}_i + \boldsymbol{v}_i \tag{5.7}$$

---

[11] Blundell and Smith (1991) consider a range of CMLE estimators, from the case without the full error components restrictions, to the fully stationary error components model (which is equivalent to the ML estimator in Bhagarva and Sargan (1983)).

in which $\tilde{\eta}_i = \eta_i - \phi y_{i1}$. From Eq. (5.6) we have $\mathrm{E}(y_{i1}\tilde{\eta}_i) = 0$ and $\mathrm{var}(\tilde{\eta}_i)$ is common across individuals. The conditional model (5.7) therefore retains an error components structure with homoskedastic errors.

Defining $\tilde{\Omega}$ to be the $(T-1) \times (T-1)$ error covariance matrix for the errors $\tilde{\boldsymbol{u}}_i = \boldsymbol{l}\tilde{\eta}_i + \boldsymbol{v}_i$ in system (5.7), we then have the following moment condition:

$$\mathrm{E}(\boldsymbol{y}'_{i,-1}\tilde{\Omega}^{-1}\tilde{\boldsymbol{u}}_i) = \boldsymbol{0}. \tag{5.8}$$

As a result error components GLS will be consistent for α in the conditional model (5.7).[12] The parameter ϕ is identified for $T \geqslant 3$ through the restrictions between A and $\boldsymbol{a}$. Under the normality assumption

$$\boldsymbol{u}_i \sim \mathrm{N}(\boldsymbol{l}\eta_i, \boldsymbol{\Omega}) \quad \text{for all } i = 1, \ldots, N \tag{5.9}$$

this CGLS estimator is equivalent to the conditional MLE. For the autoregressive error components model with homoskedastic normal disturbances and satisfying assumption (5.6), this CGLS estimator is therefore efficient for α provided there are no cross equation restrictions between α and ϕ.[13]

6. Monte Carlo results

In this section we report the results of a Monte Carlo study which investigates the potential gains from exploiting the moment conditions (4.3) and (4.4) in finite samples, as well as the potential gains from using the conditional GLS estimator.

6.1. Design

We consider two data generation processes for y_{it}.

Model A:

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

Model B:

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

---

[12] Note that it remains the case that $\mathrm{E}(\boldsymbol{y}'_{i,-1}\tilde{\boldsymbol{u}}_i) \neq \boldsymbol{0}$. The orthogonality condition (5.8) 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 $\tilde{\Omega}$. As indicated in Prucha (1984), this is a situation when the asymptotic covariance matrix for the *feasible* conditional GLS estimator of α depends on the distribution of the estimated $\tilde{\Omega}$. Bootstrap methods would, however, be applicable in this parametric autoregressive model, as described in Horowitz (1997).

[13] See Blundell and Smith (1991).

for $i = 1, 2, \ldots, N$ and $t = 2, 3, \ldots, T$, where in each case the η_i and v_{it} are drawn as mutually independent i.i.d. N(0,1) random variables. In model B, the contribution of the time-invariant component of the error term becomes less important as the autoregressive parameter α increases.

In both cases we generate the initial conditions y_{i1} as:

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

where u_{i1} is an i.i.d. N(0,σ_{u1}^2) random variable, independent of both η_i and v_{it}, and with the variance of u_{i1} chosen to satisfy stationarity. Some models with non-stationary initial conditions, and non-normal and heteroskedastic disturbances, are considered in the appendix. All results are based on 1000 Monte Carlo replications, with new values for the initial conditions drawn in each replication.

We present results for three GMM estimators:

DIF: The standard first-differenced GMM estimator, based on moment conditions (3.1) only.

SYS: The system GMM estimator, based on moment conditions (3.1), (4.3) and (4.4).

ALL: The system GMM estimator which also exploits the further $T - 2$ moment restrictions (4.8) that follow from assumption (4.4) and the homoskedasticity through time of v_{it}.

Note that the linear GMM (ALL) estimator exploits the complete set of second-order moment restrictions implied by assumptions (2.3)–(2.6), (3.3) and (4.4).

The results reported below are for conventional two-step GMM estimators. For GMM (DIF), the choice of A_N to yield an initial consistent estimator is that discussed in footnote 3. In our Monte Carlo designs, these one-step and two-step GMM (DIF) estimators are asymptotically equivalent, and their finite sample distributions were found to be similar in our simulations. For GMM (SYS) and GMM (ALL), the choice of A_N to yield an initial consistent estimator is $A_N = (N^{-1} \sum_{i=1}^{N} Z_i^+{}' Z_i^+)^{-1}$, which is asymptotically inefficient relative to the two-step estimators.[14] However the finite sample distributions of the one- and two-step system GMM estimators were found to be similar in most of our simulations. This is discussed further in the appendix. For comparison, we also report results for a feasible conditional GLS estimator, which uses residuals from the one-step GMM (SYS) estimator to estimate the required variance components.

---

[14] Without prior knowledge of, for example, $\sigma_\eta^2 / \sigma_v^2$, there is no one-step weight matrix that is asymptotically equivalent to two-step GMM in these cases.

6.2. Results

The main results are presented in Table 2. We consider sample sizes with $N = 100$, 200 and 500, and values of α varying from 0.0 to 0.9. Table 2a reports results for model A with $T = 4$, Table 2b reports results for model B with $T = 4$, and Table 2c reports further results for model A with $T = 11$.

Our main findings concern the poor performance of the first-differenced GMM estimator at high values of α, and the dramatic improvement that results from exploiting the additional moment conditions (4.3) and (4.4) in these cases.

Table 2
(a) Stationary model A, $T = 4$

N	α	Mean	GMM2 (DIF) RMSE SD	Mean	GMM2 (SYS) RMSE SD	Mean	GMM2 (ALL) RMSE SD	Mean	CGLS RMSE SD
100	0.0	− 0.0044	0.1227 *0.1227*	0.0100	0.0994 *0.0990*	0.0060	0.0970 *0.0969*	0.0157	0.0986 *0.0974*
	0.3	0.2865	0.1853 *0.1849*	0.3132	0.1221 *0.1215*	0.3100	0.1216 *0.1213*	0.3188	0.1228 *0.1215*
	0.5	0.4641	0.2693 *0.2674*	0.5100	0.1333 *0.1330*	0.5100	0.1356 *0.1353*	0.5182	0.1353 *0.1342*
	0.8	0.4844	0.8805 *0.8224*	0.8101	0.1620 *0.1618*	0.8169	0.1541 *0.1533*	0.8365	0.1396 *0.1349*
	0.9	0.2264	1.0659 *0.8264*	0.9405	0.1615 *0.1564*	0.9422	0.1415 *0.1351*	0.9572	0.1121 *0.0964*
200	0.0	− 0.0037	0.0854 *0.0854*	0.0051	0.0670 *0.0669*	0.0028	0.0651 *0.0651*	0.0083	0.0700 *0.0696*
	0.3	0.2919	0.1272 *0.1270*	0.3092	0.0838 *0.0833*	0.3061	0.0812 *0.0810*	0.3120	0.0895 *0.0887*
	0.5	0.4828	0.1828 *0.1821*	0.5098	0.0941 *0.0936*	0.5079	0.0925 *0.0922*	0.5135	0.1015 *0.1006*
	0.8	0.6362	0.5468 *0.5219*	0.8050	0.1196 *0.1195*	0.8112	0.1143 *0.1138*	0.8259	0.1115 *0.1085*
	0.9	0.3731	1.1000 *0.9661*	0.9235	0.1499 *0.1481*	0.9308	0.1243 *0.1205*	0.9431	0.1022 *0.0927*
500	0.0	− 0.0033	0.0557 *0.0556*	0.0012	0.0434 *0.0434*	0.0001	0.0421 *0.0422*	0.0025	0.0462 *0.0461*
	0.3	0.2936	0.0827 *0.0824*	0.3025	0.0552 *0.0552*	0.3008	0.0530 *0.0530*	0.3030	0.0607 *0.0606*
	0.5	0.4887	0.1177 *0.1172*	0.5021	0.0632 *0.0632*	0.5006	0.0612 *0.0612*	0.5025	0.0710 *0.0710*
	0.8	0.7386	0.3144 *0.3085*	0.7939	0.0781 *0.0779*	0.7942	0.0770 *0.0769*	0.8007	0.0853 *0.0853*
	0.9	0.5978	0.7081 *0.6407*	0.9043	0.1000 *0.0999*	0.9038	0.0884 *0.0883*	0.9172	0.0880 *0.0863*

(b) Stationary model B, *T = 4*

N	α	GMM2 (DIF)		Mean	GMM2 (SYS)		Mean	GMM2 (ALL)		Mean	CGLS
		Mean	RMSE SD			RMSE SD			RMSE SD		RMSE SD
100	0.0	− 0.0044	0.1227 *0.1227*	0.0100		0.0994 *0.0990*	0.0060		0.0970 *0.0969*	0.0157	0.0986 *0.0974*
	0.3	0.2895	0.1589 *0.1586*	0.3042		0.1134 *0.1134*	0.3018		0.1141 *0.1141*	0.3075	0.1140 *0.1138*
	0.5	0.4833	0.1901 *0.1895*	0.4939		0.1194 *0.1193*	0.4931		0.1212 *0.1211*	0.4944	0.1200 *0.1199*
	0.8	0.7586	0.2844 *0.2816*	0.7737		0.1336 *0.1311*	0.7762		0.1326 *0.1305*	0.7638	0.1294 *0.1242*
	0.9	0.8115	0.4033 *0.3936*	0.8626		0.1494 *0.1447*	0.8670		0.1426 *0.1388*	0.8505	0.1361 *0.1268*
200	0.0	− 0.0037	0.0854 *0.0854*	0.0051		0.0670 *0.0669*	0.0028		0.0651 *0.0651*	0.0083	0.0700 *0.0696*
	0.3	0.2939	0.1087 *0.1086*	0.3046		0.0775 *0.0774*	0.3027		0.0764 *0.0764*	0.3060	0.0830 *0.0828*
	0.5	0.4915	0.1280 *0.1278*	0.5003		0.0816 *0.0816*	0.4995		0.0816 *0.0817*	0.5003	0.0877 *0.0877*
	0.8	0.7808	0.1842 *0.1833*	0.7926		0.0890 *0.0887*	0.7927		0.0898 *0.0896*	0.7850	0.0920 *0.0908*
	0.9	0.8607	0.2488 *0.2458*	0.8883		0.0967 *0.0960*	0.8885		0.0967 *0.0961*	0.8768	0.0966 *0.0938*
500	0.0	− 0.0033	0.0557 *0.0556*	0.0012		0.0434 *0.0434*	0.0001		0.0421 *0.0422*	0.0025	0.0462 *0.0461*
	0.3	0.2943	0.0708 *0.0706*	0.2999		0.0513 *0.0513*	0.2991		0.0502 *0.0502*	0.3004	0.0563 *0.0564*
	0.5	0.4920	0.0831 *0.0828*	0.4969		0.0539 *0.0538*	0.4966		0.0535 *0.0534*	0.4969	0.0604 *0.0604*
	0.8	0.7851	0.1201 *0.1192*	0.7938		0.0550 *0.0547*	0.7937		0.0553 *0.0549*	0.7888	0.0615 *0.0605*
	0.9	0.8759	0.1619 *0.1602*	0.8932		0.0563 *0.0559*	0.8930		0.0567 *0.0563*	0.8850	0.0619 *0.0600*

(c) Stationary model A, *T = 11*

N	α	GMM2 (DIF)		GMM2 (SYS)		GMM2 (ALL)		CGLS	
		Mean	RMSE SD	Mean	RMSE SD	Mean	RMSE SD	Mean	RMSE SD
100	0.0	− 0.0138	0.0483 *0.0463*	− 0.0183	0.0468 *0.0431*	− 0.0153	0.0467 *0.0441*	− 0.0071	0.0364 *0.0358*
	0.3	0.2762	0.0591 *0.0541*	0.2728	0.0558 *0.0487*	0.2795	0.0545 *0.0506*	0.2832	0.0424 *0.0389*
	0.5	0.4629	0.0725 *0.0623*	0.4689	0.0618 *0.0535*	0.4794	0.0592 *0.0555*	0.4761	0.0490 *0.0428*
	0.8	0.6812	0.1576 *0.1036*	0.7925	0.0655 *0.0651*	0.8043	0.0624 *0.0623*	0.8025	0.0595 *0.0595*
	0.9	0.6455	0.2996 *0.1581*	0.9259	0.0522 *0.0453*	0.9302	0.0523 *0.0428*	0.9422	0.0623 *0.0459*
200	0.0	− 0.0070	0.0358 *0.0352*	− 0.0059	0.0310 *0.0304*	− 0.0057	0.0313 *0.0307*	− 0.0037	0.0272 *0.0270*
	0.3	0.2883	0.0427 *0.0411*	0.2914	0.0345 *0.0335*	0.2925	0.0348 *0.0340*	0.2907	0.0318 *0.0304*
	0.5	0.4815	0.0503 *0.0468*	0.4899	0.0373 *0.0359*	0.4922	0.0373 *0.0365*	0.4858	0.0369 *0.0340*
	0.8	0.7373	0.0971 *0.0742*	0.8025	0.0421 *0.0420*	0.8075	0.0430 *0.0423*	0.8039	0.0449 *0.0448*
	0.9	0.7256	0.2152 *0.1261*	0.9231	0.0435 *0.0369*	0.9263	0.0445 *0.0359*	0.9345	0.0506 *0.0370*
500	0.0	− 0.0025	0.0201 *0.0200*	− 0.0010	0.0172 *0.0172*	− 0.0012	0.0173 *0.0173*	− 0.0016	0.0169 *0.0168*
	0.3	0.2959	0.0237 *0.0233*	0.2986	0.0182 *0.0181*	0.2984	0.0183 *0.0182*	0.2960	0.0196 *0.0192*
	0.5	0.4934	0.0276 *0.0268*	0.4984	0.0189 *0.0189*	0.4983	0.0190 *0.0190*	0.4937	0.0228 *0.0220*
	0.8	0.7695	0.0536 *0.0441*	0.8019	0.0244 *0.0243*	0.8027	0.0249 *0.0248*	0.8008	0.0316 *0.0316*
	0.9	0.8110	0.1168 *0.0757*	0.9120	0.0306 *0.0280*	0.9135	0.0312 *0.0282*	0.9206	0.0361 *0.0296*

This is particularly the case for model A, where the permanent effects remain important even at high values of α, and when only a small number of time-series observations are available.

Consider first the experiments where α is 0.8 or 0.9 in Table 2a. For GMM (DIF) we find both a huge downward bias and very imprecise estimates. This is consistent with our analysis in Section 3.2, where we showed that there was a serious problem of weak instruments for the GMM (DIF) estimator at values of α around 0.8 and above, particularly in model A. In contrast, we find both a much smaller bias and much improved precision for either of the system

GMM estimators. For example, at $\alpha = 0.8$ and $N = 200$ the root mean squared error (RMSE) for GMM (SYS) is only 20% of the RMSE for GMM (DIF). These gains from using conditions (4.3) and (4.4) remain very impressive, even at $N = 500$.

Consistent with our theoretical analysis, the finite-sample bias of the first-differenced GMM estimator is less serious in model B, as shown in Table 2b. The weak instruments problem depends on the persistence in the level of the series, which decreases both with α *and* with the importance of the permanent effects. Even in this case, though, there remain substantial efficiency gains from exploiting conditions (4.3) and (4.4). For example, at $\alpha = 0.8$ and $N = 200$ the RMSE for GMM (SYS) is still 50% of the RMSE for GMM (DIF).

Similarly the poor performance of GMM (DIF) improves with the number of time periods available, as shown for model A in Table 2c. Nevertheless, for high values of α there remain impressive gains in bias, precision and RMSE, even at $T = 11$. For example, at $\alpha = 0.8$ and $N = 200$ the RMSE for GMM (SYS) remains less than half of the RMSE for GMM (DIF).

The gains from exploiting the homoskedasticity restrictions (4.8) are found to be much smaller. The biggest gains are found for highly persistent series and low values of T. At $\alpha = 0.9$ and $T = 4$, the RMSE for GMM (ALL) is around 85% of the RMSE for GMM (SYS) in model A. However at $T = 11$ we find that there is no gain from exploiting the homoskedasticity restrictions, and the same is found in model B even at $T = 4$.

Conditional GLS also offers much better performance than the first-differenced GMM estimator. At $N = 500$, the variance of conditional GLS is generally found to be higher than the variance of the system GMM estimators.[15] However at $N = 100$ the finite sample performance of conditional GLS is generally found to be better than that of the system GMM estimators.[16]

Table 3 summarises the gains in precision found in Table 2, in the same format as the asymptotic efficiency comparisons presented in Ahn and Schmidt (1995, Table 1) and used in Table 1 in Section 4.4. The figures in Table 3 show the ratio of the empirical variance of the two-step GMM (DIF) estimator in our simulations to the empirical variance of the two-step GMM (SYS) estimator. High values of this ratio imply large efficiency gains when the additional moment conditions (4.3) and (4.4) are exploited.[17] As would be expected from the discussion of Table 1, the efficiency gains are largest at high values of α, low values of T, and at higher values of the variance of the permanent effects (relative to the variance of v_{it}). The empirical variance ratios found at $N = 500$ in

---

[15] Exceptions are found for $\alpha = 0.9$ in Table 2a and for $\alpha = 0.0$ in Table 2c.

[16] At $T = 4$ this is particularly found at high values of α, although this is not repeated when $T = 11$.

[17] Notice that by neglecting the reductions in bias, these ratios tend to underestimate the improvement in root mean squared error.

Table 3
Empirical variance ratios

N	α	(a) $T = 4$ Model A $\dfrac{\mathrm{var}(GMM - DIF)}{\mathrm{var}(GMM - SYS)}$	(b) $T = 4$ Model B $\dfrac{\mathrm{var}(GMM - DIF)}{\mathrm{var}(GMM - SYS)}$	(c) $T = 11$ Model B $\dfrac{\mathrm{var}(GMM - DIF)}{\mathrm{var}(GMM - SYS)}$
100	0.0	1.54	1.54	1.15
	0.3	2.32	1.96	1.23
	0.5	4.04	2.52	1.36
	0.8	25.84	4.61	2.53
	0.9	27.92	7.40	12.18
200	0.0	1.63	1.63	1.34
	0.3	2.32	1.97	1.51
	0.5	3.79	2.45	1.70
	0.8	19.07	4.27	3.12
	0.9	42.55	6.56	11.68
500	0.0	1.64	1.64	1.35
	0.3	2.29	1.89	1.66
	0.5	3.44	2.37	2.01
	0.8	14.91	4.75	3.29
	0.9	41.13	8.21	7.31

Table 3a are reasonably close to the corresponding asymptotic variance ratios given in the $T = 4$, $\sigma_\eta^2/\sigma_v^2 = 1$ results of Table 1.

Further Monte Carlo results presented in the appendix investigate the reliability of inference based on the usual asymptotic variance formulae for the GMM estimators. The main result is that whilst asymptotic t-tests based on the one-step GMM estimators are found to have the correct empirical level in our simulations, the asymptotic t-tests based on the two-step GMM estimators can be seriously misleading, and tend to reject too frequently. This tendency is exaggerated when the errors are non-normal or heteroskedastic.

7. An application to employment equations

7.1. The model

In this section we illustrate the benefits of exploiting the additional linear moment restrictions (4.3) and (4.4) in an application using real data. We consider a simple dynamic labour demand equation of the form

$$n_{it} = \alpha n_{i,t-1} + \beta_0 w_{it} + \beta_1 w_{i,t-1} + \gamma_0 k_{it} + \gamma_1 k_{i,t-1} + (\eta_i + v_{it}),$$

where n_{it} is the log of employment in firm i in year t, w_{it} is the log of the wage rate and k_{it} is the log of the capital stock. This log-linear model can be derived as the static conditional labour demand equation for a competitive firm with Cobb–Douglas technology, with the dynamics resulting from an AR(1) disturbance as in model (2.2). However, we do not impose the implied common factor restrictions, and alternatively the dynamics may be thought of as an empirical approximation to some more general adjustment process.

In the context of model (2.1) 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 Eqs. (4.3) and (4.4) 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 (4.4) on the initial conditions process this allows the use of both lagged Δy_{it} and lagged Δx_{it} as instruments in levels equations.

As we do not expect wages and capital to be strictly exogenous in our employment application, we focus here on moment conditions of the type $E(u_{it}\Delta y_{i,t-1}) = 0$ and $E(u_{it}\Delta x_{i,t-1}) = 0$. Stricter exogeneity restrictions on the x_{it} would allow the use of further instruments, as detailed in Arellano and Bond (1991), Ahn and Schmidt (1995) and Arellano and Bover (1995). For the same reason we do not consider conditional GLS results here. Extension of conditional GLS to models with regressors requires strict exogeneity, as discussed in Blundell and Smith (1991).

7.2. Data

The data we use is taken from Arellano and Bond (1991). Briefly, this is an unbalanced panel of 140 UK listed manufacturing companies with between 7 and 9 annual observations over the period 1976–1984. The data starts in 1976 only because Datastream did not report employment in earlier years. Hence, there is nothing special about the first observation on the firms in this sample, and we might expect the initial conditions restriction (4.4) to be valid here.

7.3. Results for a short sample period

We first consider omitting the first 3 observations for each firm, leaving an unbalanced panel with between 4 and 6 observations between 1979 and 1984. The first two columns of Table 4 report the results for GMM (DIF) and GMM (SYS) respectively.[18] These are the one-step GMM estimators, for which we

---

[18] Both wages and capital are treated as potentially endogenous variables, so that for GMM (DIF) the instrument set contains observations on (n, w, k) dated t-2 and earlier. For GMM (SYS) the instruments used in the levels equations are the observations on $(\Delta n, \Delta w, \Delta k)$ dated t-1. The results are obtained using DPD98 for GAUSS, which can be obtained by email from steve.bond@nuffield.ox.ac.uk.

Table 4
Employment equations

	1979–84 GMM-DIF	1979–84 GMM-SYS	1976–84 GMM-DIF	1976–84 GMM-SYS
$n_{i,t-1}$	0.4046	0.8596	0.7075	0.8103
	(0.1363)	(0.0680)	(0.0842)	(0.0578)
w_{it}	− 1.1073	− 1.0255	− 0.7088	− 0.7968
	(0.3888)	(0.1807)	(0.1171)	(0.1002)
$w_{i,t-1}$	0.3947	0.7705	0.5000	0.5488
	(0.1643)	(0.1527)	(0.1113)	(0.1488)
k_{it}	0.5463	0.5625	0.4660	0.4268
	(0.2571)	(0.2229)	(0.1010)	(0.0771)
$k_{i,t-1}$	− 0.4216	− 0.4769	− 0.2151	− 0.2786
	(0.2821)	(0.2420)	(0.0859)	(0.0784)
$m1$	− 2.24	− 3.94	− 5.60	− 6.50
$m2$	1.13	1.04	− 0.14	− 0.08
Sargan	15.34	43.43	88.80	116.05
(df)	(25)	(37)	(79)	(100)

Notes.
1. Year dummies are included in all specifications.
2. Asymptotic standard errors, asymptotically robust to heteroskedasticity, are reported in parentheses.
3. m1 and m2 are tests for first-order and second-order serial correlation in the first-differenced residuals, asymptotically distributed as $N(0,1)$ under the null of no serial correlation.
4. Sargan is a test of the over-identifying restrictions, asymptotically distributed as χ^2 under the null of instrument validity, with degrees of freedom reported in parentheses.
5. The instruments used in each equation are
GMM (DIF) – $n_{i,t-2}, n_{i,t-3}, \ldots, n_{i1}$; $w_{i,t-2}, w_{i,t-3}, \ldots, w_{i1}$; $k_{i,t-2}, k_{i,t-3}, \ldots, k_{i1}$.
GMM (SYS) – differenced equations – as GMM (DIF); levels equations – $\Delta n_{i,t-1}$, $\Delta w_{i,t-1}$, $\Delta k_{i,t-1}$.

believe inference based on the asymptotic variance matrix to be more reliable.[19] Using the first-differenced GMM estimator in this short panel, the coefficient on the lagged dependent variable is only 0.4, and the elasticity of employment with respect to capital is only 0.2, suggesting implausibly low returns to scale. Using the system GMM estimator which exploits the moment conditions (4.3) and (4.4), the coefficient on the lagged dependent variable is 0.86, and the elasticity of employment with respect to capital is a more plausible 0.6. We also find a substantial improvement in precision, and the extended set of moment restrictions is not rejected by the Sargan/Hansen test of over-identifying

---

[19] See Section 6.2 and the appendix for further discussion. The reported standard errors are asymptotically robust to heteroskedasticity.

restrictions. The tests of serial correlation in the first-differenced residuals are in both cases consistent with the maintained assumption of no serial correlation in v_{it}.

7.4. Results for a longer sample period

The last two columns of Table 4 report the results for the same estimators, but using the full sample period from 1976 to 1984 as originally used in the Arellano and Bond (1991) study. The main difference is that the GMM (DIF) estimates are now much closer to the GMM (SYS) estimates: the coefficient on the lagged dependent variable is now 0.71, and the elasticity of employment with respect to capital is found to be around 0.8 for both estimators. There remains a gain in precision from exploiting the additional moment restrictions, and their validity is not rejected in this example.

The results of this section appear to be consistent with our earlier analysis. With short sample periods and persistent series, we find important small sample biases and imprecision using the standard first-differenced GMM estimator. These biases can be substantially reduced, and more precise parameter estimates can be obtained, if restriction (4.4) is satisfied and we exploit the extra moment conditions that this implies.

8. 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 extend naturally 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 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. The second approach conditions on the observed initial values to obtain a system that under certain conditions can be estimated consistently by error components GLS.

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 the system GMM estimator can be considerably more efficient than non-linear GMM in this case.

Acknowledgements

Appendix A.

In this appendix we present some further Monte Carlo results, which consider designs with non-stationary initial conditions, non-normal and heteroskedastic disturbances; and we investigate the reliability of inference based on the usual asymptotic variance formulae for the GMM estimators. We focus on the case with $T = 4$, $N = 200$, $\alpha = 0.5$, $\sigma_\eta^2 = 1$ and $\sigma_v^2 = 1$.

Table 5 reproduces the base case results for this configuration from Table 2a. Here we present the one-step GMM estimators as well as the two-step GMM estimators, and we present OLS levels, within groups and standard error components GLS results for comparison. As expected in this i.i.d. design, the one-step GMM(DIF) estimator performs as well as the two-step GMM(DIF) estimator. More interestingly, the loss in efficiency from using the one-step GMM(SYS) and GMM(ALL) estimators is quite modest.

For the GMM estimators, we also report the number of rejections (from 1000 replications) of a Wald test of the null hypothesis that $\alpha = 0.5$, and of the standard Sargan/Hansen test of over-identifying restrictions. The one-step Wald tests appear to have the correct empirical level, but the two-step Wald tests tend to reject too frequently, particularly for the system estimators where there are more over-identifying restrictions. This is primarily because the asymptotic standard errors for the two-step GMM estimators tend to underestimate the true dispersion of these estimators in finite samples. For example, the mean asymptotic standard error (ASE) of the two-step GMM(ALL) estimator is only 89% of its empirical standard deviation.

Table 6 reports an experiment with non-stationary initial conditions, in which the initial values are mean zero noise. This results in serious biases for the system GMM estimators. The conventional Sargan test has power to detect these biases.

Table 7 presents a stationary design but with non-normal errors. The v_{it} disturbances are drawn from a skewed $\chi^2(1)$ distribution, then centred and

Table 5
Stationary, normal, homoskedastic base case

Estimator	Mean	Std. Dev.	Mean ASE	RMSE
OLS	0.8745	0.0185		0.3749
Within	− 0.0346	0.0467		0.5366
GLS	0.6684	0.0961		0.1944
GMM1(DIF)	0.4809	0.1783	0.1822	0.1792
GMM2(DIF)	0.4828	0.1821	0.1798	0.1828
GMM1(SYS)	0.5040	0.1079	0.1104	0.1079
GMM2(SYS)	0.5098	0.0936	0.0892	0.0941
GMM1(ALL)	0.5118	0.1073	0.1106	0.1079
GMM2(ALL)	0.5079	0.0922	0.0820	0.0925
CGLS	0.5135	0.1006		0.1015

Rejection frequencies

Estimator	One-Step Wald			Two-Step Wald			Two-Step Sargan		
	10%	5%	1%	10%	5%	1%	10%	5%	1%
DIF	83	49	15	97	64	18	110	54	8
SYS	99	56	11	153	95	34	97	59	11
ALL	100	56	14	167	106	43	98	50	13

$y_{i1} = 2\eta_i + u_{i1}$ with $u_{i1} \sim$ i.i.d. N(0,4/3).
$\sigma_\eta^2 = 1$, $\sigma_v^2 = 1$, $\alpha = 0.5$, $T = 4$, $N = 200$.

Table 6
Non-stationary initial conditions

Estimator	Mean	Std. Dev.	Mean ASE	RMSE
OLS	0.7000	0.0331		0.2028
Within	0.2852	0.0356		0.2177
GLS	0.4997	0.0396		0.0396
GMM1(DIF)	0.4972	0.0555	0.0567	0.0555
GMM2(DIF)	0.4972	0.0563	0.0561	0.0564
GMM1(SYS)	0.7006	0.0726	0.0602	0.2133
GMM2(SYS)	0.6866	0.1041	0.0542	0.2137
GMM1(ALL)	0.7032	0.0759	0.0600	0.2169
GMM2(ALL)	0.6852	0.1017	0.0523	0.2113
CGLS	0.5020	0.0482		0.0482

Rejection Frequencies

Estimator	One-Step Wald			Two-Step Wald			Two-Step Sargan		
	10%	5%	1%	10%	5%	1%	10%	5%	1%
DIF	91	49	15	98	52	18	110	55	7
SYS	927	883	728	832	791	672	1000	1000	1000
ALL	923	875	731	838	801	679	1000	1000	1000

$y_{i1} \sim$ i.i.d. N(0,16/3).
$\sigma_\eta^2 = 1$, $\sigma_v^2 = 1$, $\alpha = 0.5$, $T = 4$, $N = 200$.

Table 7
Non-normal errors

Estimator	Mean	Std Dev	Mean ASE	RMSE
OLS	0.8740	0.0203		0.3746
Within	− 0.0343	0.0565		0.5373
GLS	0.6659	0.0965		0.1919
GMM1(DIF)	0.4867	0.1844	0.1775	0.1848
GMM2(DIF)	0.4911	0.1840	0.1700	0.1841
GMM1(SYS)	0.4999	0.1082	0.1068	0.1081
GMM2(SYS)	0.5032	0.0938	0.0825	0.0938
GMM1(ALL)	0.5067	0.1109	0.1078	0.1111
GMM2(ALL)	0.5045	0.0981	0.0762	0.0981
CGLS	0.5124	0.1030		0.1037

Rejection frequencies

Estimator	One-Step Wald			Two-Step Wald			Two-Step Sargan		
	10%	5%	1%	10%	5%	1%	10%	5%	1%
DIF	103	55	10	124	65	12	94	44	8
SYS	105	59	19	171	102	41	84	45	4
ALL	108	70	25	228	155	75	116	55	8

$v_{it} = (\varepsilon_{it} - 1)/\sqrt{2}$ where $\varepsilon_{it} \sim \chi^2(1)$.
$\sigma_\eta^2 = 1$, $\sigma_v^2 = 1$, $\alpha = 0.5$, $T = 4$, $N = 200$.

Table 8
Heteroskedastic errors across individuals

Estimator	Mean	Std Dev	Mean ASE	RMSE
OLS	0.8748	0.0228		0.3754
Within	− 0.0351	0.0537		0.5378
GLS	0.6644	0.1117		0.1988
GMM1(DIF)	0.4585	0.2410	0.2368	0.2444
GMM2(DIF)	0.4715	0.2388	0.2253	0.2404
GMM1(SYS)	0.5011	0.1248	0.1258	0.1248
GMM2(SYS)	0.5127	0.1116	0.0976	0.1122
GMM1(ALL)	0.5145	0.1254	0.1266	0.1262
GMM2(ALL)	0.5148	0.1076	0.0868	0.1086
CGLS	0.5114	0.1161		0.1166

Rejection frequencies

Estimator	One-Step Wald			Two-Step Wald			Two-Step Sargan		
	10%	5%	1%	10%	5%	1%	10%	5%	1%
DIF	114	64	19	127	71	25	114	56	5
SYS	118	62	18	194	125	58	101	48	4
ALL	118	66	25	214	156	69	114	50	10

$v_{it} \sim N(0,\sigma_i^2)$ where $\sigma_i^2 = 0.5 + 0.5(y_{i1}^2/\text{var}(y_{i1}))$.
$\sigma_\eta^2 = 1$, $\alpha = 0.5$, $T = 4$, $N = 200$.

Table 9
Heteroskedastic errors over time

Estimator	Mean	Std Dev	Mean ASE	RMSE
OLS	0.8998	0.0199		0.4003
Within	0.1214	0.0683		0.3798
GLS	0.7502	0.0755		0.2613
GMM1(DIF)	0.4478	0.2165	0.2111	0.2226
GMM2(DIF)	0.4636	0.1952	0.1894	0.1985
GMM1(SYS)	0.5023	0.1157	0.1158	0.1157
GMM2(SYS)	0.5182	0.0804	0.0754	0.0824
GMM1(ALL)	0.4313	0.1199	0.1117	0.1382
GMM2(ALL)	0.5854	0.0999	0.0732	0.1314
CGLS	0.6110	0.0903		0.1430

Rejection Frequencies

Estimator	One-Step Wald			Two-Step Wald			Two-Step Sargan		
	10%	5%	1%	10%	5%	1%	10%	5%	1%
DIF	102	41	6	95	54	11	108	54	10
SYS	102	55	20	148	78	28	103	47	10
ALL	157	85	11	403	311	180	1000	1000	1000

$v_{it} \sim N(0, \sigma_t^2)$ where $\sigma_2^2 = 0.2$, $\sigma_3^2 = 1$ and $\sigma_4^2 = 1.8$.
$\sigma_\eta^2 = 1$, $\alpha = 0.5$, $T = 4$, $N = 200$.

scaled to have mean zero and variance one. Comparing Tables 5 and 7 suggest that this non-normality has little impact on the means and standard deviations of these estimators. However, there is a further deterioration in the quality of inference based on the asymptotic standard errors for the two-step GMM estimators. The mean ASE for the two-step GMM(ALL) estimator falls to just 78% of its empirical standard deviation in this design.

Table 8 presents a design with heteroskedasticity across individuals, in which the variance of v_{it} is dependent on y_{i1}^2. These disturbances are again scaled such that they have variance one across the whole sample. This form of heteroskedasticity considerably increases the variance of the GMM(DIF) estimators, but has a smaller impact on the variance of the system GMM estimators. There is a small improvement in the precision of the two-step GMM estimators relative to the one-step GMM estimators. More striking again is the downward bias in the (heteroskedasticity consistent) asymptotic standard errors for the two-step GMM estimators. Inference based on the one-step GMM estimators appears to be much more reliable when either non-normality or heteroskedasticity is suspected.

Finally, Table 9 presents a design with heteroskedasticity over time. As expected, this results in serious biases for both the GMM(ALL) estimator and for conditional GLS. In the former case, the conventional Sargan test has power to detect the invalidity of the overidentifying restrictions in this design.

References

Ahn, S.C., Schmidt, P., 1995. Efficient estimation of models for dynamic panel data. Journal of Econometrics 68, 5–28.

Alonso-Borrego, C., Arellano, M., 1996. Symmetrically normalised instrumental variable estimation using panel data. CEMFI Working Paper No. 9612, September.

Anderson, T.W., Hsiao, C., 1981. Estimation of dynamic models with error components. Journal of the American Statistical Association 76, 598–606.

Arellano, M., Bond, S.R., 1991. Some tests of specification for panel data: Monte Carlo evidence and an application to employment equations. Review of Economic Studies 58, 277–297.

Arellano, M., Bover, O., 1995. Another look at the instrumental-variable estimation of error-components models. Journal of Econometrics 68, 29–52.

Bhagarva, A., Sargan, J.D., 1983. Estimating dynamic random effects models from panel data covering short time periods. Econometrica 51, 1635–1659.

Blundell, R.W., Smith, R.J., 1991. Initial conditions and efficient estimation in dynamic panel data models. Annales d'Economie et de Statistique 20/21, 109–123.

Chamberlain, G., 1987. Asymptotic efficiency in estimation with conditional moment restrictions. Journal of Econometrics 34, 305–334.

Hansen, L.P., 1982. Large sample properties of generalised method of moment estimators. Econometrica 50, 1029–1054.

Holtz-Eakin, D., Newey, W., Rosen, H.S., 1988. Estimating vector autoregressions with panel data. Econometrica 56, 1371–1396.

Horowitz, J.L., 1997. Bootstrap methods in econometrics: theory and numerical performance, in: Kreps, D.M., Wallis, K.F. (Eds.), Advances in Economics and Econometrics: Thoery and Applications vol. III, Chapter 8, Econometric Society Monograph No. 28. Cambridge University Press, Cambridge.

Hsiao, C., 1986, Analysis of Panel Data. Cambridge University Press, Cambridge.

Nelson, C.R., Startz, R., 1990a. Some further results on the exact small sample properties of the instrumental variable estimator. Econometrica 58, 967–976.

Nelson, C.R., Startz, R., 1990b. The distribution of the instrumental variable estimator and its t-ratio when the instrument is a poor one. Journal of Business Economics and Statistics 63, 5125–5140.

Nickell, S.J., 1981. Biases in dynamic models with fixed effects. Econometrica 49, 1417–1426.

Prucha, I.R., 1984. On the asymptotic efficiency of feasible aitken estimators for seemingly unrelated regression models with error components. Econometrica 52, 203–207.

Sevestre, P., Trognon, A., 1985. A note on autoregressive error components models. Journal of Econometrics 29, 231–245.

Staiger, D., Stock, J.H., 1997. Instrumental variables regression with weak instruments. Econometrica 65, 557–586.