ANNALES D'ÉCONOMIE ET DE STATISTIQUE. – N° 55-56 – 1999

Estimating a Dynamic Panel Data Model with Heterogenous Trends

Tom J. Wansbeek, Thijs Knaap *

ABSTRACT. – This paper is concerned with the estimation of a dynamic panel data model with individual fixed effects and a linear trend term with heterogenous coefficients. We expand the available methods for the standard dynamic panel data model to this case, and discuss modified OLS, IV and GMM methods. We hint at the potential of LIML estimation to reduce small sample bias. Although the methods pertain to a more complex model than the standard dynamic model, the simplicity of the various approaches can also be brought to bear on the standard model.

Estimation d'un modèle dynamique pour des données individuelles temporelles avec des tendances hétérogènes

RÉSUMÉ. – On considère ici l'estimation d'un modèle dynamique pour des données individuelles temporelles avec des effets fixes et une tendance linéaire hétérogène. Nous étendons les méthodes disponibles pour le cas standard sans tendance (OLS, IV et GMM) et suggérons le potentiel de LIML pour réduire le biais dans des échantillons petits.

* T.J. WANSBEEK, T. KNAAP, Department of Economics, University of Groningen, The Netherlands. We are indebted to Lászlo MÁTYÁS and two referees for their useful comments on an earlier version of this paper.

1 Introduction

This paper is concerned with the estimation of a dynamic panel data model with individual fixed effects and a linear trend term with heterogenous coefficients. We expand the available methods for the standard dynamic panel data model with only individual effects to this case, and discuss modified OLS, IV and GMM methods. We hint at the potential of LIML estimation to reduce small sample bias. Although the methods pertain to a more complex model than the standard dynamic model, the simplicity of the various approaches can also be brought to bear on the standard model.

In section 2, we introduce the model under consideration. In section 3 we consider OLS estimation, and in section 3 we discuss IV estimation. Since the justification of IV techniques is asymptotic (in the cross-sectional dimension) and data sets *e.g.* pertaining to countries are necessarily small in this sense, small-sample properties of estimators require special attention. Section 3 contains a brief discussion of a LIML approach, which could possess better small-sample properties than the IV estimator.

The IV estimators exploit only a part of the set of orthogonality conditions induced by the model. Efficiency could be enhanced by extending the estimation procedure to take all orthogonality conditions into account. This calls for a GMM estimator, which is derived in section 4. In section 4a we present some simulation results suggesting the performance of the various methods.

2 The Model

The model to be analyzed in the sequel is the first-order autoregressive model with a constant and a deterministic trend:

$$z_{i,t} = \gamma z_{i,t-1} + \mu_i + \xi_i t + \varepsilon_{i,t} \tag{1}$$

This is a panel data model with an individual effect, and with varying coefficients attached to a time trend. There is a common coefficient for the lagged endogenous variable. We assume stationarity, letting $-1 < \gamma < 1$. The model is relevant for a number of fields in economics, for instance the empirical analysis of the SOLOW [1956] model. A version of the SOLOW model (with all ξ_i equal) predicts that a country's income per capita will follow the process (1).

It is useful to rewrite the model (1) in matrix format. Define

$$\begin{aligned} z_i &\equiv (z_{i1}, \ldots, z_{iT})' \\ z_{i,-1} &\equiv (z_{i0}, \ldots, z_{i,T-1})' \\ f &\equiv (1, \ldots, T)' \end{aligned}$$

$$\iota_T \equiv (1,\ldots,1)'$$
$$e_1 \equiv (1,0,\ldots,0)'$$
$$\varepsilon_i \equiv (\varepsilon_{i1},\ldots,\varepsilon_{iT})'.$$

Then the model in matrix format per observation is

$$z_i = \gamma z_{i,-1} + \xi_i f + \mu_i \iota_T + \varepsilon_i,$$

with $i = 1,\ldots,N$ indexing cross-sectional entities.

In this paper we consider the case where both sets of effects are fixed. In order to eliminate the effects we take second differences. First-differencing eliminates the individual effect μ_i. The time trend f is then transformed into ι_{T-1}, again creating an individual effect. First-differencing again eliminates this on its turn. Using Δ to denote first differences, the implied model is

$$\Delta\Delta z_i = \gamma \Delta\Delta z_{i,-1} + \Delta\Delta\varepsilon_i, \tag{2}$$

where the various vectors are of order $T-3$, so

$$w_i \equiv \Delta\Delta z_i = \begin{bmatrix} z_{i4} - 2z_{i3} + z_{i2} \\ \vdots \\ z_{it} - 2z_{i,t-1} + z_{i,t-2} \\ \vdots \\ z_{iT} - 2z_{i,T-1} + z_{i,T-2} \end{bmatrix}, \tag{3}$$

and $w_{i,-1} \equiv \Delta\Delta z_{i,-1}$ and $u_i \equiv \Delta\Delta\varepsilon_i$ are defined analogously. The model then simply is

$$w_i = \gamma w_{i,-1} + u_i. \tag{4}$$

We now discuss a variety of estimation methods for this model. Throughout, we consider asymptotics in the sense of large N and fixed T.

3 OLS Estimation

The first estimation method to be considered is OLS in (4). This estimator is defined by

$$\hat{\gamma} \equiv \frac{\sum_i w_{i,-1}' w_i}{\sum_i w_{i,-1}' w_{i,-1}}. \tag{5}$$

OLS in this model yields evidently an inconsistent estimator of γ due to the correlation that the differencing operations induced between the regressor and the disturbance term.

It can still be worthwhile to consider OLS, for the following reason. The plim of the OLS estimator when $N \to \infty$ is a function of γ. When this function is of a simple form the inverse function of the estimator is a straightforward (and possibly robust) consistent estimator of γ. This is the case for a dynamic panel data model with individual effects only. If these are eliminated by taking first differences, the plim ($N \to \infty$) of the OLS estimator is $(\gamma - 1)/2$. This indicates a severe bias but $2\hat{\gamma}_{\text{OLS}} + 1$ is by construction consistent, and extremely simple to compute.

In the case of twice differencing the situation is slightly more complex. In order to evaluate the plim of the OLS estimator we use the following result. Consider in general the model $z_i = \gamma z_{i,-1} + \varepsilon_i$, with ε i.i.d., and let Q be symmetric p.s.d. Let

$$\hat{\gamma} \equiv \frac{\sum_i z'_{i,-1} Q z_i}{\sum_i z'_{i,-1} Q z_{i,-1}}.$$

Then

$$\operatorname*{plim}_{N\to\infty} \hat{\gamma} = \gamma + \frac{1-\gamma^2}{2\gamma}\left(1 - \frac{\operatorname{tr} Q}{\operatorname{tr} Q\Psi_\gamma}\right), \tag{6}$$

where Ψ_γ is the AR(1) correlation matrix with autocorrelation parameters γ:

$$\Psi_\gamma = \begin{bmatrix} 1 & \gamma & & \gamma^{T-1} \\ \gamma & \ddots & \ddots & \\ & \ddots & \ddots & \gamma \\ \gamma^{T-1} & & \gamma & 1 \end{bmatrix}.$$

This expression (see also WANSBEEK and KNAAP, [1998]) generalizes well-known expressions from the literature for two leading cases. In the first place, for a dynamic model with fixed individual effects, the plim of the "within" estimator results when Q is taken equal to the centering operator of order T: $Q = I_T - \iota_T \iota'_T / T$. Then an expression first derived by NICKELL [1981] results. It can also be used to derive the plim of the OLS estimator after taking first differences. Then, letting D_T denote the $T \times (T-1)$ matrix that takes first differences, $Q = D_T D'_T$ and the above mentioned result $(\gamma - 1)/2$ is obtained, probably first derived by CHAMBERLAIN [1980].

We can use (6) also in the present context. Then the plim of the OLSE in (4) follows from (6) by substituting

$$Q = D_T D_{T-1} D'_{T-1} D'_T.$$

After some algebra we obtain

$$\begin{aligned} \operatorname{tr} Q &= 6(T-2) \\ \operatorname{tr} Q\Psi_\gamma &= 2(1-\gamma)(3-\gamma)(T-2), \end{aligned}$$

and

$$\operatorname*{plim}_{N\to\infty} \hat{\gamma} = \frac{-\gamma^2 + 3\gamma - 4}{2(3-\gamma)}. \tag{7}$$

Applying the inverse function gives

$$\tilde{\gamma} \equiv \hat{\gamma}_{\mathrm{OLS}} + \frac{3}{2} - \sqrt{\left(\hat{\gamma}_{\mathrm{OLS}} - \frac{7}{2}\right)\left(\hat{\gamma}_{\mathrm{OLS}} + \frac{1}{2}\right)} \qquad (8)$$

as a consistent estimator for γ.

This procedure is simple but has two important defects. In the first place (6) holds under the assumption of stationarity. That is, it is assumed that $|\gamma| < 1$, and the process has been going on for an infinite amount of time before the observation period starts. The absence of a unit root is an assumption that we wish to avoid here. A second, problematic and somewhat surprising drawback is that (7) only yields values outside the interval (– 1/2, 7/2). In particular, $0 \leqslant \gamma \leqslant 1$ corresponds with $-2/3 \leqslant \hat{\gamma} \leqslant -1/2$; and there is no guarantee whatsoever that $\hat{\gamma}$ satisfies these bounds. These two problems make the modified OLS approach to consistent estimation useless for our purpose, so we have to consider more complicated approaches to estimation.

4 IV Estimators

Since $\mathrm{E} z_{i\tau} \Delta\Delta\varepsilon_{i,t} = 0$ for $\tau \leqslant t-3$ we can generalize the approach to estimate dynamic panel data models by instrumental variables due to ANDERSON and HSIAO [1981, 1982] to the present case. In fact, the only difference is that we have a twice-differenced instead of a once-differenced model. We obtain

$$\hat{\gamma}_{t,\tau} \equiv \frac{\sum_i z_{i\tau} \Delta\Delta z_{it}}{\sum_i z_{i\tau} \Delta\Delta z_{i,t-1}} \qquad (9)$$

as consistent estimators for γ for any choice of $t, 4 \leqslant t \leqslant T$ and $1 \leqslant \tau \leqslant t-3$.

So we have a variety of consistent estimators of γ that will differ in finite samples. An obvious way to avoid the non-uniqueness is to pool resources (cf. ARELLANO and BOND [1991]) and derive the instrumental variables estimator found by combining all orthogonality conditions. Let

$$
\begin{aligned}
m &\equiv \frac{(T-2)(T-3)}{2} \\
Z_i &\equiv \begin{bmatrix} z_{i1} & & & & \\ & z_{i1} & & & \\ & z_{i2} & & & \\ & & \ddots & & \\ & & & z_{i1} & \\ & & & \vdots & \\ & & & z_{i,T-3} & \end{bmatrix} \\
Z' &\equiv (Z_1, \ldots, Z_n) \\
P_Z &\equiv Z(Z'Z)^{-1}Z' \\
w &\equiv \begin{bmatrix} w_1 \\ \vdots \\ w_N \end{bmatrix}
\end{aligned} \tag{10}
$$

and w_{-1} equals w after lagging all elements by one period. The various matrices and vectors are of order: $Z_i : m \times (T-3), Z'$: $m \times N(T-3)$, $w_i : (T-3) \times 1, w : N(T-3) \times 1$. So

$$\mathrm{E} Z_i \Delta\Delta\varepsilon_i = \mathrm{E} Z_i (w_i - \gamma w_{i,-1}) = 0.$$

This orthogonality condition can be employed to derive a number of IV estimators. The first one of these amounts to OLS after transforming the model (2) by multiplying by the IV's. This yields

$$\hat{\gamma}_{\mathrm{IV1}} \equiv \frac{\left(\sum_i Z_i w_{i,-1}\right)' \left(\sum_i Z_i w_i\right)}{\left(\sum_i Z_i w_{i,-1}\right)' \left(\sum_i Z_i w_{i,-1}\right)} = \frac{w'_{-1} Z Z' w}{w'_{-1} Z Z' w_{-1}}. \tag{11}$$

The second IV estimator has the usual projection form,

$$\hat{\gamma}_{\mathrm{IV2}} \equiv \frac{\left(\sum_i Z_i w_{i,-1}\right)' \left(\sum_i Z_i Z_i'\right)^{-1} \left(\sum_i Z_i w_i\right)}{\left(\sum_i Z_i w_{i,-1}\right)' \left(\sum_i Z_i Z_i'\right)^{-1} \left(\sum_i Z_i w_{i,-1}\right)} = \frac{w'_{-1} P_Z w}{w'_{-1} P_Z w_{-1}}. \tag{12}$$

These estimators are both less satisfactory since the double data transformation has induced a non-scalar covariance matrix of the disturbances. We may obtain a better estimator if we correct for this structure. Let

$$\mathrm{E}\left(u_i u_i'\right) = \sigma_\varepsilon^2 \Theta$$

with

$$\Theta = \begin{bmatrix} 6 & -4 & 1 & & \\ -4 & \ddots & \ddots \ddots & & \\ 1 & \ddots & \ddots & \ddots & 1 \\ & \ddots & \ddots & \ddots & -4 \\ & & 1 & -4 & 6 \end{bmatrix}.$$

Then $\Theta^{-\frac{1}{2}} u_i$ is homoscedastic. Because of

$$\mathrm{E}(Z_i \Theta^{\frac{1}{2}})(\Theta^{-\frac{1}{2}} u_i) = 0,$$

we can adapt the above by replacing Z_i by $Z_i \Theta^{\frac{1}{2}}$, w_i by $\Theta^{-\frac{1}{2}} w_i$, etc., leading to the estimator

$$\hat{\gamma}_{\mathrm{IV3}} \equiv \frac{\left(\sum_i Z_i w_{i,-1}\right)' \left(\sum_i Z_i \Theta Z_i'\right)^{-1} \left(\sum_i Z_i w_i\right)}{\left(\sum_i Z_i w_{i,-1}\right)' \left(\sum_i Z_i \Theta Z_i'\right)^{-1} \left(\sum_i Z_i w_{i,-1}\right)} =$$

$$\frac{w_{-1}' Z (Z' \bar{\Theta} Z)^{-1} Z' w}{w_{-1}' Z (Z' \bar{\Theta} Z)^{-1} Z' w_{-1}}, \tag{13}$$

with $\bar{\Theta} \equiv I_n \otimes \Theta$. The asymptotic variances of the estimators are given by

$$\mathrm{avar} \hat{\gamma}_{\mathrm{IV1}} = \sigma_\varepsilon^2 \mathrm{plim}\ N \frac{w_{-1}' Z Z' \bar{\Theta} Z Z' w_{-1}}{(w_{-1}' Z Z' w_{-1})^2}$$

$$\mathrm{avar} \hat{\gamma}_{\mathrm{IV2}} = \sigma_\varepsilon^2 \mathrm{plim}\ N \frac{w_{-1}' P_Z \bar{\Theta} P_Z w_{-1}}{(w_{-1}' P_Z w_{-1})^2}$$

$$\mathrm{avar} \hat{\gamma}_{\mathrm{IV3}} = \sigma_\varepsilon^2 \mathrm{plim}\ N \left(w_{-1}' Z (Z' \bar{\Theta} Z)^{-1} Z' w_{-1} \right)^{-1},$$

where the asymptotics is of the usual “large N” variety.

5 LIML Estimation

The estimators introduced up till now (and the GMM estimators to be discussed later on) derive their justification from asymptotic arguments. They have desirable properties for $N \to \infty$. It is well-known that their behavior in small samples may be less than satisfactory and in particular can exhibit a bias, see *e.g.* ALTONJI and SEGAL [1996]. If we think for example of the growth literature, which pertains to countries, whose number is not large anyhow, methods that have better small-sample properties are desirable. In this section we discuss the LIML estimator, which is one method with potential in this respect.

To introduce the LIML estimator it is useful to reconsider the (second) IV estimator. It can be considered as the minimizer of

$$q = (w - \gamma w_{-1})' P_Z (w - \gamma w_{-1}).$$

The LIML estimator can correspondingly be introduced as the minimizer of

$$
(14) \quad \begin{aligned} q^* &= \frac{(w - \gamma w_{-1})' P_Z (w - \gamma w_{-1})}{(w - \gamma w_{-1})'(I - P_Z)(w - \gamma w_{-1})} \\ &= \frac{(1,-\gamma)(w,w_{-1})' P_Z (w,w_{-1})(1,-\gamma)'}{(1,-\gamma)(w,w_{-1})'(I - P_Z)(w,w_{-1})(1,-\gamma)'}, \end{aligned}
$$

where the identity matrix is of order $n(T-3)$. The first-order condition is

$$
\left((w,w_{-1})' P_Z (w,w_{-1}) - \lambda (w,w_{-1})'(I - P_Z)(w,w_{-1})\right) \begin{bmatrix} -1 \\ \gamma \end{bmatrix} = 0.
$$

So the LIML estimator follows as the solution of an eigenvalue equation involving matrices of order 2, where the smallest eigenvalue and the corresponding eigenvector is appropriate in order to minimize q^*. Following the ideas due to BEKKER [1994], the LIML estimator could have superior small-sample properties as compared to the IV estimator. In particular, LIML estimators are consistent under a wider variety of asymptotics than IV or 2SLS. LIML estimation is, notwithstanding its long standing, seldom used and has as far as we know not been applied to panel data contexts. Again, we can adapt the estimator for the induced correlation over time. The adapted estimator is the minimizer of

$$
(15) \quad q^{**} = \frac{(1,-\gamma)(w,w_{-1})' \bar{P}_Z (w,w_{-1})(1,-\gamma)'}{(1,-\gamma)(w,w_{-1})'(\bar{\Theta}^{-1} - \bar{P}_Z)(w,w_{-1})(1,-\gamma)'},
$$

where $\bar{P}_Z \equiv Z(Z'\bar{\Theta}Z)^{-1}Z'$.

Somewhat adapting from BEKKER (1994), the intuition behind LIML (properly, pseudo-LIML, since no assumption of normality has been made) can be described as follows. Consider the regression equation $y = X\beta + \varepsilon$ and let $\mathrm{E}(\varepsilon|Z) = 0$. The projection of the IV's Z on X is

$$
Z = X\Pi + V,
$$

say. Let $u \equiv \varepsilon + V\beta$. Then

$$
(y,X) = Z\Pi(\beta,I) + (u,V)
$$

If the rows of (u,V) are i.i.d. with variance Ω, then

$$
(16) \quad \Sigma \equiv \mathrm{E}\left\{(y,X)' P_Z (y,X)\right\} = (\beta,I)'\Pi' Z' Z\Pi(\beta,I) + l\Omega
$$

$$
(17) \quad \Sigma_\perp \equiv \mathrm{E}\left\{(y,X)'(I - P_Z)(y,X)\right\} = (n-l)\Omega,
$$

with l the number of instruments and n the number of observations. Let $\lambda \equiv l/(n-l)$. Notice that

$$
(18) \quad (\beta,I)\begin{bmatrix} 1 \\ -\beta \end{bmatrix} = 0.
$$

Substituting (17) in (16) gives

$$
\Sigma = (\beta,I)'\Pi' Z' Z\Pi(\beta,I) + \lambda\Sigma_\perp,
$$

and rearranging and using (18) gives

$$(\Sigma - \lambda\Sigma_\perp)\begin{bmatrix} 1 \\ -\beta \end{bmatrix} = 0. \tag{19}$$

Since Σ and $\Sigma_\perp$ are not observed but their sample counterparts S and $S_\perp$ are, we can turn the identity (19) into an estimating equation by defining the LIML estimator $\hat{\beta}_{\text{LIML}}$ as the solution of the eigenequation

$$(S - \lambda S_\perp)\begin{bmatrix} 1 \\ -\hat{\beta}_{\text{LIML}} \end{bmatrix} = 0 \tag{20}$$

corresponding with the smallest eigenvalue. Much econometric work is based on large-sample asymptotics. In the present notation, λ is implicitly taken to be zero, and $S_\perp$ drops out of the computation. This is justified when the sample is truly large or when the instruments are not weak. Otherwise, the estimator defined by (20) "stays closer to the data" and may hence have good properties when the sample is small or the instruments are weak.

Application of LIML to the dynamic panel data model is not straightforward. The expectations (16) and (17) are not easily evaluated in the case the IV's are lagged dependent variables, hence the expectations can hold approximately at best. At present, the asymptotic distribution of the above mentioned LIML estimator is unknown.

6 An Optimal GMM Estimator

As AHN and SCHMIDT [1995, 1997] and WANSBEEK and BEKKER [1996] have shown, the IV approaches described in section 4 are inefficient since they do not employ all available moment conditions; and the gain in efficiency by using all moment conditions can be sizeable. Hence we will derive all such conditions, and present the optimal GMM estimator implied. Since this estimator is based on nonlinear moment conditions, a two-step procedure suggests itself where an initial consistent IV estimator is improved in a second step based on linearization in order to obtain an efficient estimator.

We derive the moment conditions as follows. Rewriting the model slightly, letting

$$z_{i,-1} = Dz_i + e_1 z_{0i},$$

gives

$$z_i - \gamma D z_i = \gamma z_{i0} e_1 + \xi_i f + \mu_i \iota_T + \varepsilon_i.$$

We take the individual unobservables z_{i0},ξ_i and μ_i as fixed and require the transformation that projects the corresponding regressors out. Let $X \equiv (e_1, \iota_T, f)$. Let $M \equiv I_T - X(X'X)^{-1}X'$. Then the transformed model without the individual unobservables is

$$M(z_i - \gamma D z_i) = M\varepsilon_i.$$

Under the assumption that $\mathrm{E}(\varepsilon_i \varepsilon_i') = \sigma_\varepsilon^2 I_T$, then the implied second-order structure is

$$\mathrm{E}\mathrm{vec} M \varepsilon_i \varepsilon_i' M = (\mathrm{vec} M)\sigma_\varepsilon^2. \tag{21}$$

In order to turn this into a moment condition, *i.e.* an expression involving observables and parameters with expectation zero, we project out $\mathrm{vec}M$ in order to get rid of the nuisance parameter. Since $(\mathrm{vec}M)'(\mathrm{vec}M) = \mathrm{tr}M = \mathrm{rank}M = T-3$, the projector orthogonal to $\mathrm{vec}M$ is

$$P \equiv I_{T^2} - \frac{(\mathrm{vec}M)(\mathrm{vec}M)'}{T-3}.$$

So

$$\mathrm{E}P\mathrm{vec}M\varepsilon_i \varepsilon_i' M = 0.$$

Letting

$$u_i(\gamma) \equiv M(z_i - \gamma D z_i),$$

the moment conditions implied by the model are given by the elements of

$$\mathrm{E}P(u_i(\gamma) \otimes u_i(\gamma)) = 0 \tag{22}$$

or

$$\mathrm{E}\left\{u_i(\gamma) \otimes u_i(\gamma) - \frac{u_i(\gamma)' u_i(\gamma)}{T-3}\mathrm{vec}M\right\} = 0.$$

This is a vector of T^2 elements, each containing a single moment condition for γ. The actual number of moment conditions is smaller. First, this vector is the vectorized version of a symmetric matrix so nearly half of the elements is redundant. Second, the projection P introduces an exact linear relationship between the moment conditions. Third, the same is caused by the projection M. As a result, the null space of the moment conditions vector is $(I_{T^2} - K_T, \mathrm{vec}M, (X \otimes X)(I_9 - K_3))$, with K_T the symmetric commutation matrix of order T^2, *e.g.*, MAGNUS and NEUDECKER [1988] or WANSBEEK [1989]. This null space has rank $T(T+1)/2 + 7$, hence the number of non-redundant moment conditions equals $T(T+1)/2 - 7$. So $T \geqslant 4$ is required of the data.

The GMM estimator of γ resulting from (22) is obtained by considering its sample counterpart. Write (22) in short-hand form as

$$\mathrm{E}a_i(\gamma) = 0,$$

and let

$$\bar{a}(\gamma) \equiv \frac{1}{N}\sum_i a_i(\gamma)$$

$$A(\gamma) \equiv \frac{1}{N}\sum_i a_i(\gamma) a_i(\gamma)',$$

then the GMM estimator is defined as

$$\hat{\gamma}_{\text{GMM}} = \underset{\gamma}{\text{argmin}}\, \bar{a}(\gamma)' A(\tilde{\gamma})^{+} \bar{a}(\gamma), \tag{23}$$

where $\tilde{\gamma}$ is some initial consistent estimator of γ. The first-order condition is nonlinear in γ. In particular, it is a third-order polynomial equation. The solution can be obtained straightforwardly by simple numerical methods, *e.g.* grid search over the relevant range of values of γ. The resulting estimator satisfies

$$\sqrt{N}(\hat{\gamma}_{\text{GMM}} - \gamma) \to N\left(0, (\operatorname*{plim}_{N\to\infty} \bar{a}_{\gamma}(\gamma)' A(\gamma)^{+} \bar{a}_{\gamma}(\gamma))^{-1}\right),$$

with

$$\bar{a}_{\gamma}(\gamma) \equiv \frac{\partial \bar{a}(\gamma)}{\partial \gamma}.$$

There holds

$$\frac{\partial a_i(\gamma)}{\partial \gamma} = (I_{T^2} + K_T) \left(-M(z_i - \gamma D z_i) \otimes M D z_i + \frac{2(z_i - \gamma D z_i)' M D z_i}{T-3} \text{vec} M\right).$$

In the context of the expression for the variance, the latter can be put equal to

$$\frac{\partial a_i(\gamma)}{\partial \gamma} = 2\left(-(z_i - \gamma D z_i) \otimes D z_i + \frac{1}{T-3}(z_i - \gamma D z_i)' M D z_i \text{vec} M\right).$$

In order to avoid the numerical complications involved with the solution of the first-order condition, one can employ a simpler method, that is, use an initial consistent estimator and make one step towards optimality, *cf.* NEWEY [1985].

$$\hat{\gamma}_{\text{LGMM}} = \tilde{\gamma} - \frac{\bar{a}'_{\gamma}(\tilde{\gamma}) A(\tilde{\gamma})^{+} \bar{a}(\tilde{\gamma})}{\bar{a}'_{\gamma}(\tilde{\gamma}) A(\tilde{\gamma})^{+} \bar{a}_{\gamma}(\tilde{\gamma})}.$$

This linearized GMM procedure produces an estimator with the same asymptotic properties as the non-linear procedure.

The approach is flexible in the sense that more general structures on the disturbance term can be easily accommodated. The right hand side of (21) then has a higher dimensionality and P can be easily adapted.

A number of possible adaptations of the GMM approach should be mentioned. Following ALTONJI and SEGAL [1996], the small sample behavior of the GMM estimator can be unsatisfactory. In particular, a bias in finite samples may result from the correlation of common random elements in the residual vector $(\bar{\gamma})$ and the weight matrix $A(\tilde{\gamma})$. ALTONJI and SEGAL show, for a very simple model, how this induces a downward bias. They suggest some methods to avoid this bias. One is to split the sample, and use one part of the data for the estimation of the weight matrix, and the remainder for the estimation of the residuals. This option may not always be attractive, especially in situations where the amount of data is constrained in one dimension. A better alternative then is to sacrifice some asymptotic efficiency and substitute the unit matrix for $A(\tilde{\gamma})$ so that

$$\hat{\gamma}_{\text{UGMM}} = \underset{\gamma}{\operatorname{argmin}}\, \bar{a}(\gamma)'\bar{a}(\gamma). \tag{24}$$

This unweighted GMM estimator eliminates randomness from the weight matrix and hence also eliminates any bias that this may cause. In the simulations provided by ALTONJI and SEGAL this unweighted least squares approach show good behavior.

7 Some Simulation Results

In order to give a feeling for the performance of some of the methods discussed so far, we performed a variety of simulation exercises. Because our interest lies in the small sample properties of the different estimators, we present the distributions of five estimators when $T = 7$, $T = 10$, and $T = 15$ and $N = 20$, $N = 50$, $N = 100$, and $N = 200$. All simulations use the following parameter values: $\gamma = .95$, $\sigma_\varepsilon^2 = .1$, $\xi_i \sim N(0,.0025)$, $z_{i,0} \sim N(0,9)$, and $\mu_i \sim N(0,4)$. The number of replications is 3000. Since the simulations appear to be highly time-consuming with increasing N and T, three of the twelve (N, T) combinations have not been considered in the simulations. The overall trend in the behavior of the estimators is clear for the combinations considered, so the loss in insight can be deemed limited. We have also economized somewhat on the estimators considered in the simulations, and have only made computations for the five most interesting ones: OLS, IV3, LIML2, unweighted GMM (UGMM), and weighted GMM (WGMM). The results are presented both in table 1 and figures 1 through 5. Both the table and the figures show the deviation of the estimate from the true value, $\hat{\gamma} - \gamma$. The table presents the means and standard deviations of those biases, and the figures present the distribution of the biases in simple histograms; Note that the horizontal scale differs between graphs.

In Figure 1, we see that OLS is very strongly biased, as was to be expected from the results in section 3. Compared to the other estimators, the distribution of this estimator is strongly peaked. However, as discussed in section 3, we cannot modify this estimator (and possibly retain the small variance) into a consistent estimator. Even in this simulation, most values of $\hat{\gamma}_{\text{OLS}}$ are outside the domain of formula (8). Finally, the OLS estimator is the only estimator that delivers a distribution reminiscent of the normal distribution in these small samples.

The IV3 estimator behaves remarkably well, as can be seen in Figure 2. In the smallest sample the estimator starts with a serious downward bias. This bias is retained throughout, but disappears almost after a doubling of the sample in the T dimension. A tenfold increase of the sample in the N dimension reduces the bias by the same amount, but leads to a smaller decrease in estimator variance.

TABLE 1
Statistics from the Monte Carlo Experiment

		$N = 20$		$N = 50$		$N = 100$		$N = 200$	
		$\bar{x}$	s	$\bar{x}$	s	$\bar{x}$	s	$\bar{x}$	s
	OLS	– 1.374	0.09	– 1.378	0.06	– 1.379	0.04	– 1.380	0.03
	LIML2	0.118	19.04	0.053	1.53	0.023	0.25	0.015	0.15
$T = 7$	IV3	– 0.476	0.31	– 0.248	0.23	– 0.137	0.17	– 0.067	0.12
	UGMM	– 0.935	0.50	– 0.949	0.44	– 0.964	0.40	– 0.964	0.38
	WGMM	– 0.738	0.53	– 0.779	0.46	– 0.809	0.38	– 0.862	0.39
	OLS	– 1.386	0.07	– 1.388	0.04	– 1.389	0.03		
	LIML2	0.201	14.01	0.001	0.12	0.005	0.07		
$T = 10$	IV3	– 0.196	0.14	– 0.088	0.08	– 0.041	0.06		
	UGMM	– 0.902	0.41	– 0.928	0.36	– 0.939	0.31		
	WGMM	– 0.543	0.54	– 0.514	0.53	– 0.588	0.53		
	OLS	– 1.397	0.05	– 1.400	0.03				
	LIML2	– 0.049	1.636	– 0.001	0.03				
$T = 15$	IV3	– 0.067	0.05	– 0.027	0.03				
	UGMM	– 0.912	0.48	– 0.930	0.39				
	WGMM	– 0.451	0.50	– 0.473	0.66				

$\bar{x}$ is the sample mean and s is the sample standard deviation. The estimators are in the following formulas: OLS(5), LIML2 (15), IV3 (13), UGMM (24), and GMM (23).

Figure 3 shows the LIML2 estimator, which is a close competitor of the IV3 estimator. In very small samples, this estimator exhibits a very high variance. However, this variance quickly decreases to levels comparable to those of the IV3 estimator, while the bias of the LIML2 is much less. This suggests the LIML2 as the superior estimator for $T \geqslant 10$ and $N \geqslant 50$.

The behavior of the GMM estimators is strikingly inadequate (note the horizontal scale in Figures 4 and 5). The estimator labeled UGMM, based on unweighted use of all orthogonality conditions as derived in section 6, has a downward bias that seems to *increase* as N gets larger. The fact that the variance (slowly) decreases does not help the estimator very much in this respect.

WGMM, the GMM estimator that weighs the orthogonality conditions with a matrix that is based on another estimate of γ (we used $\hat{\gamma}_{\mathrm{IV3}}$ here) has the twin-peaked distribution well-known from the weak-instruments literature (*e.g.*, NELSON and STARTZ, [1990a, 1990b]). The rightmost peak seems to converge to the true value, but the mass in the left peak ruins the performance in terms of average bias and variance.

To the extent that these simulations are indicative (a variety of other simulations with different parameter values spawned qualitatively the same results), the weighted LIML estimator corresponding with (15) comes out as superior for most practical sample sizes. We notice, though, that an investigation of the behavior of this estimator is still a topic for further research. Before it can be employed in practical situations, the first- and second-order properties of this estimator (for the case under study, but equally so for all other panel data models) should be investigated based on the alternative asymptotics as discussed by BEKKER [1994].

FIGURE 1

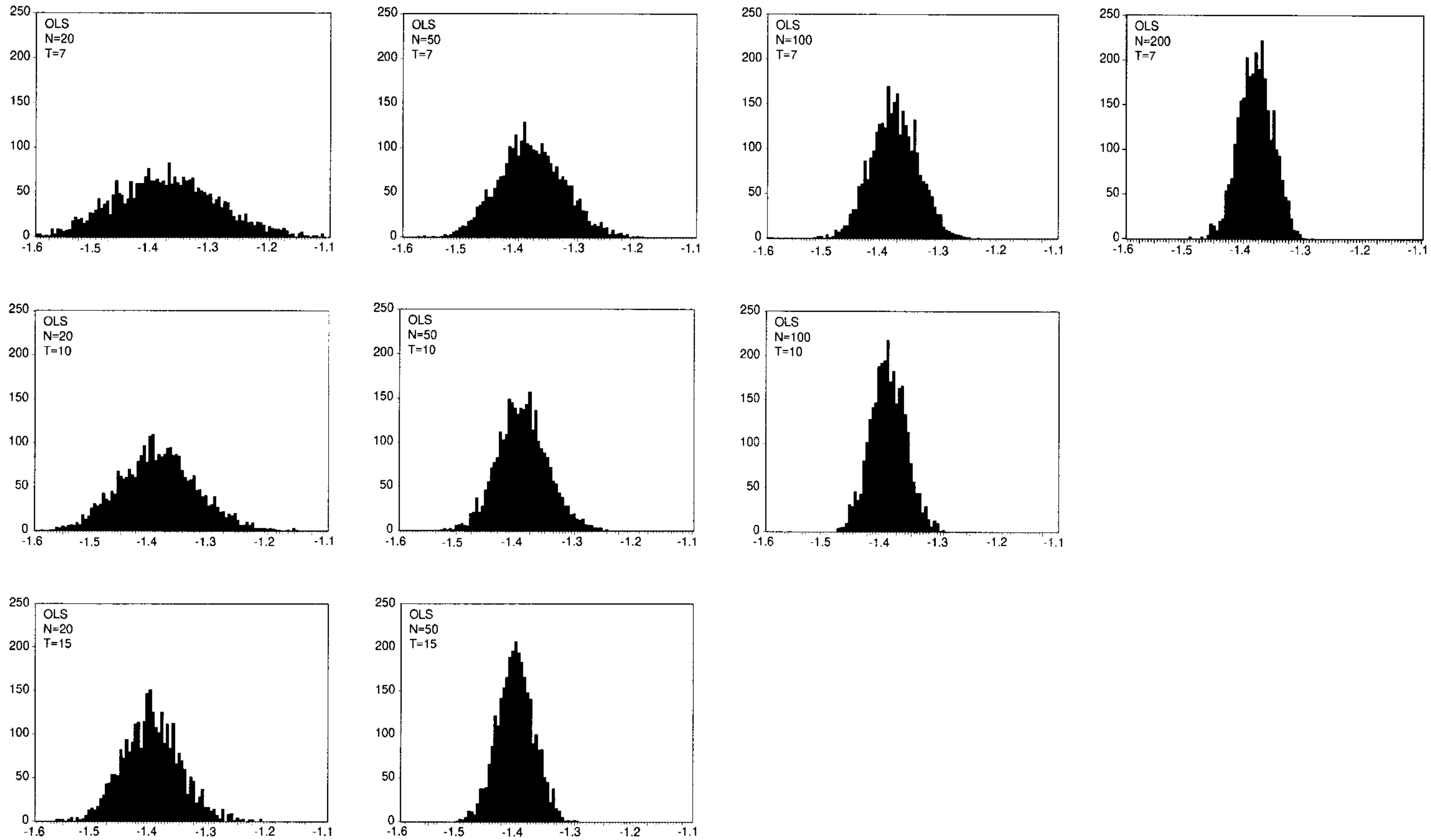

OLS
N=20
T=7
OLS
N=50
T=7
OLS
N=100
T=7
OLS
N=200
T=7
OLS
N=20
T=10
OLS
N=50
T=10
OLS
N=100
T=10
OLS
N=20
T=15
OLS
N=50
T=15
0
50
100
150
200
250
-1.6
-1.5
-1.4
-1.3
-1.2
-1.1

FIGURE 2

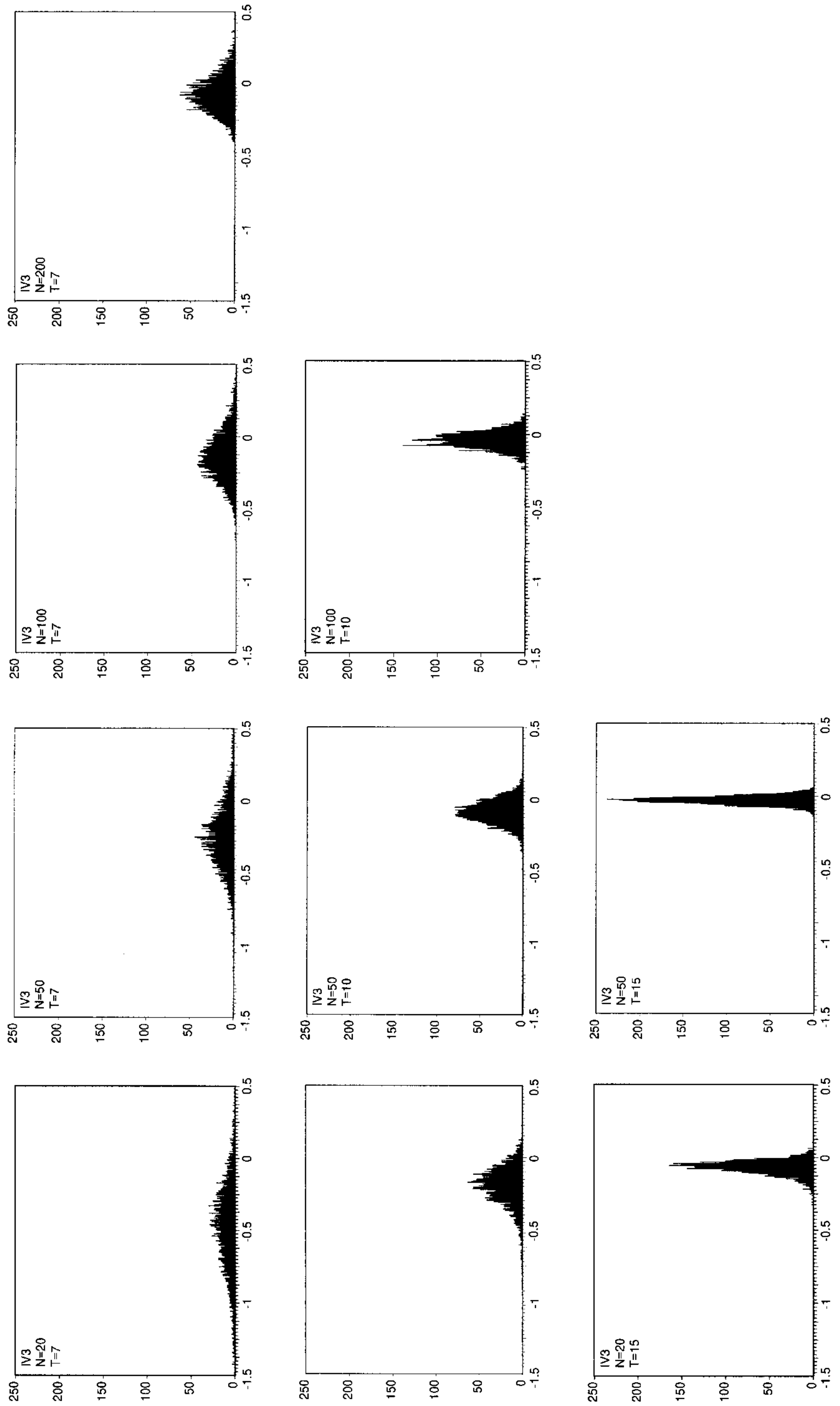

IV3 N=20 T=7
IV3 N=50 T=7
IV3 N=100 T=7
IV3 N=200 T=7
IV3 N=50 T=10
IV3 N=100 T=10
IV3 N=20 T=15
IV3 N=50 T=15

FIGURE 3

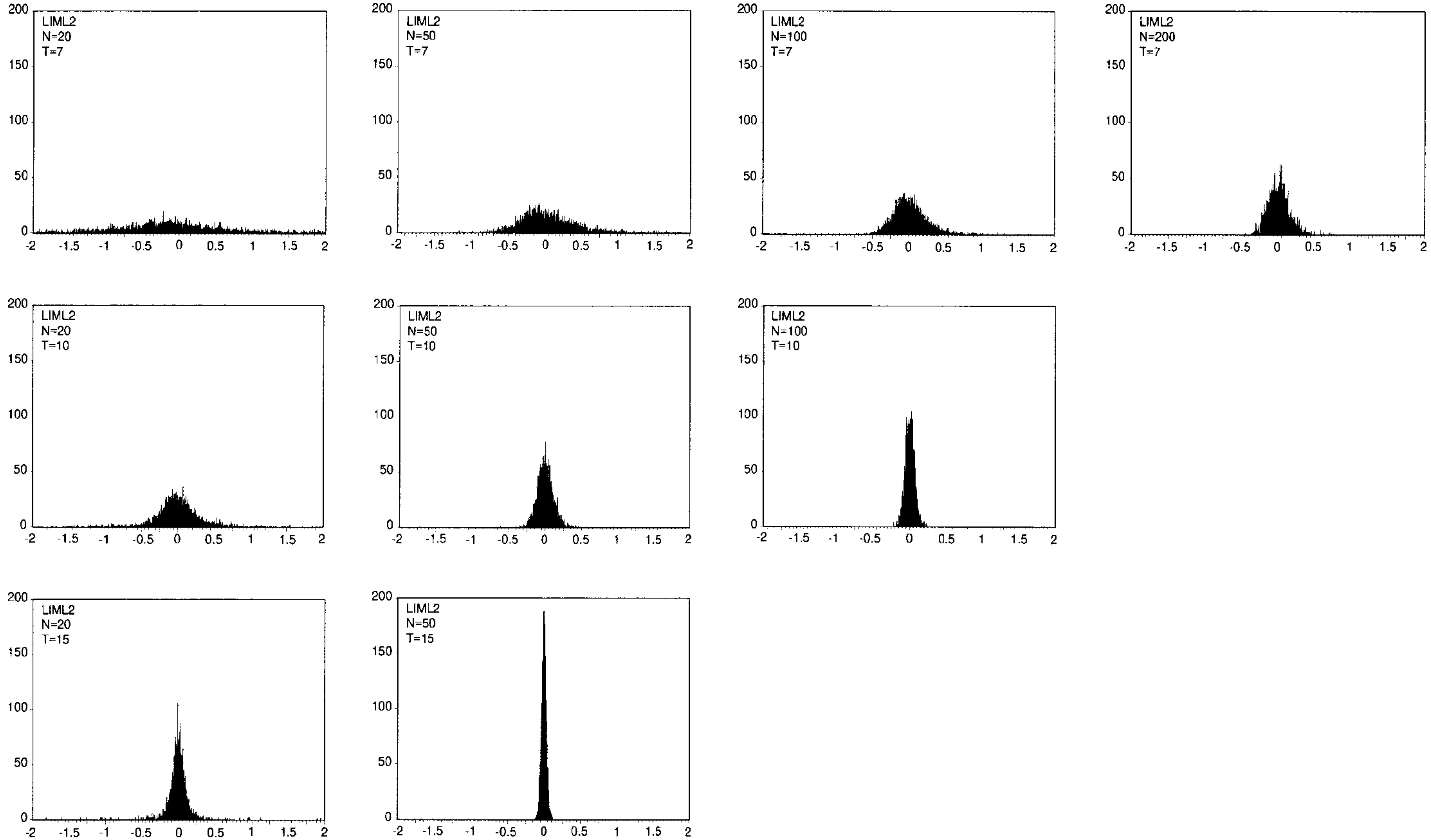

LIML2
N=20
T=7
LIML2
N=50
T=7
LIML2
N=100
T=7
LIML2
N=200
T=7
LIML2
N=20
T=10
LIML2
N=50
T=10
LIML2
N=100
T=10
LIML2
N=20
T=15
LIML2
N=50
T=15

FIGURE 4

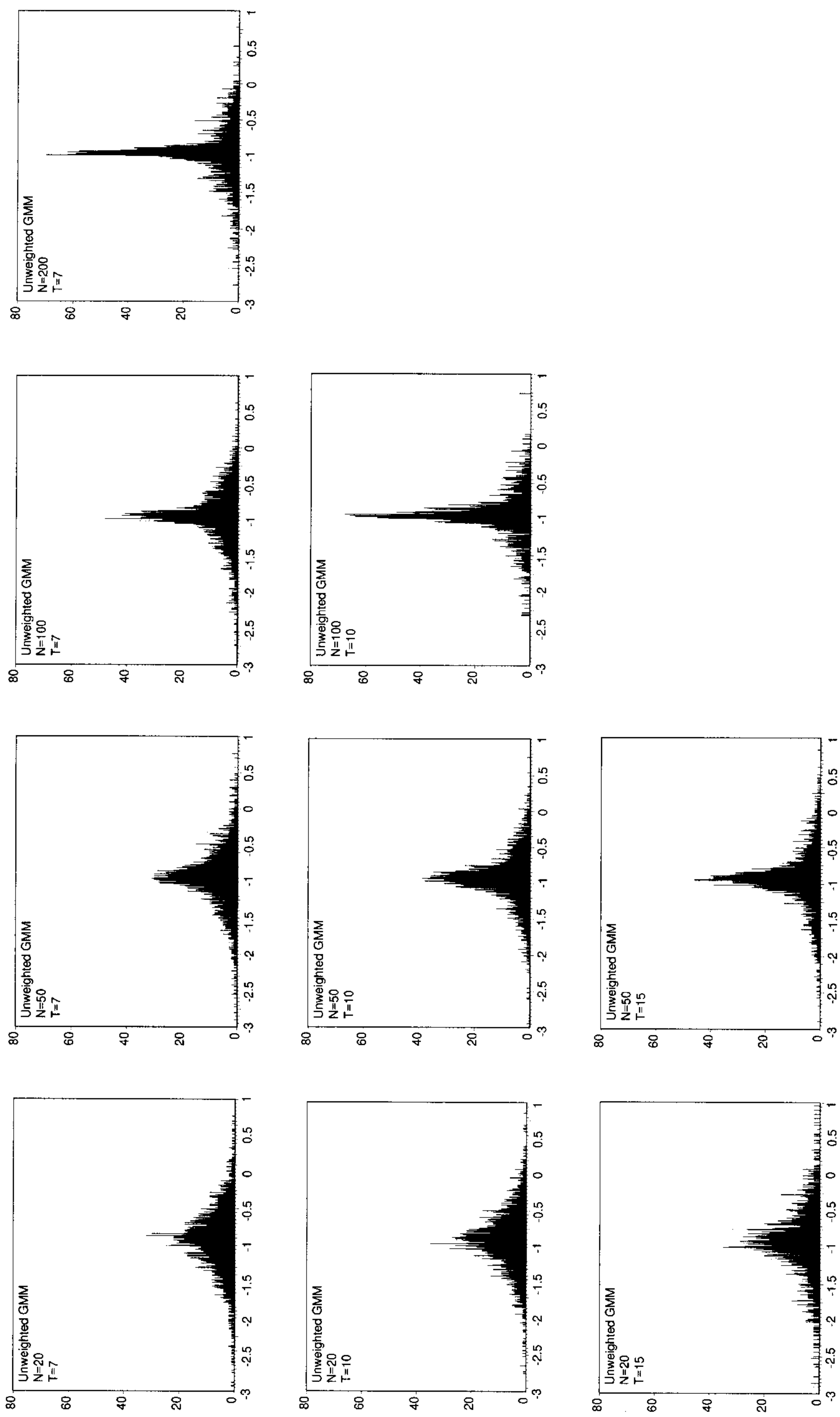
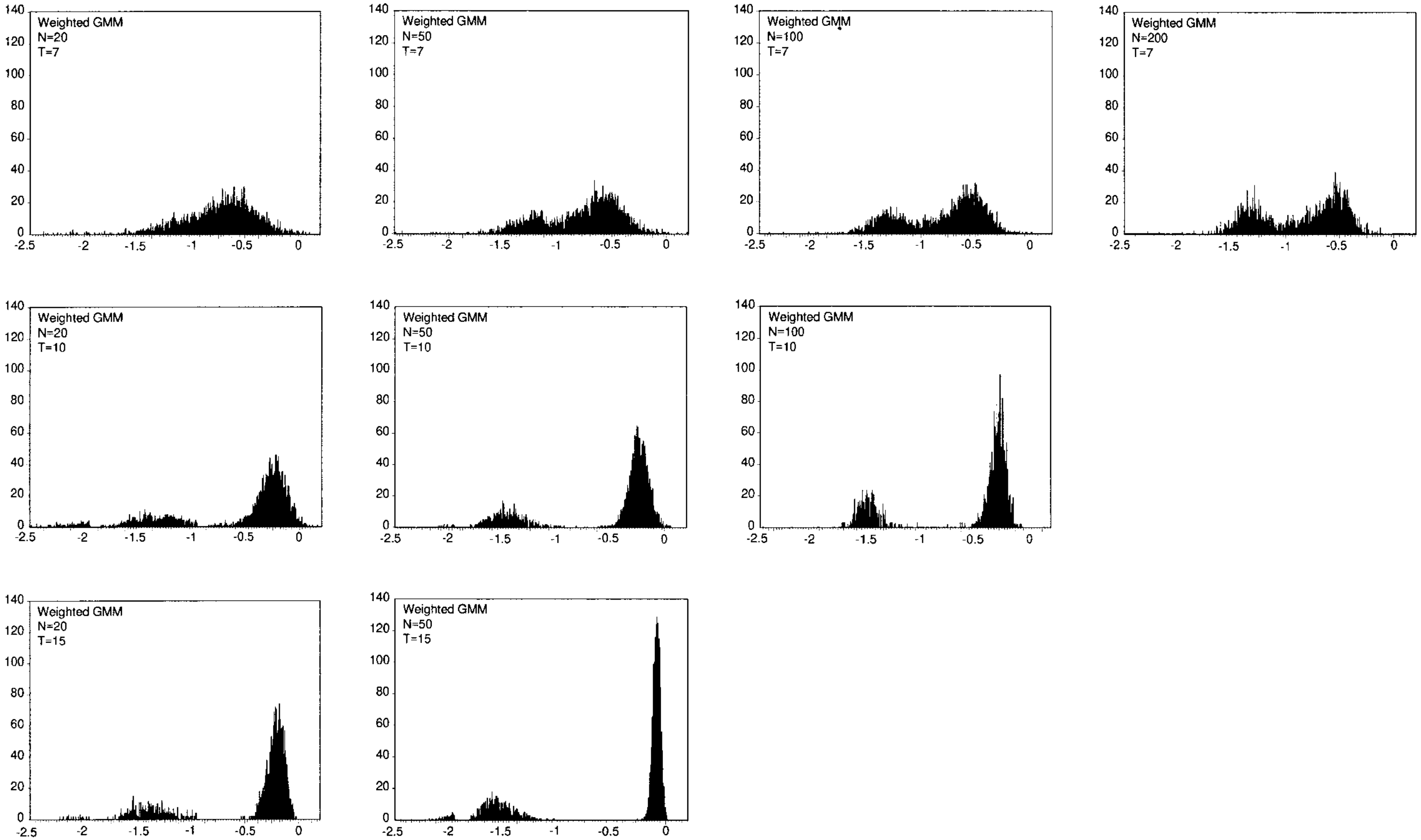

Unweighted GMM
N=20
T=7
Unweighted GMM
N=50
T=7
Unweighted GMM
N=100
T=7
Unweighted GMM
N=200
T=7
Unweighted GMM
N=20
T=10
Unweighted GMM
N=50
T=10
Unweighted GMM
N=100
T=10
Unweighted GMM
N=20
T=15
Unweighted GMM
N=50
T=15

FIGURE 5


Weighted GMM
N=20
T=7
Weighted GMM
N=50
T=7
Weighted GMM
N=100
T=7
Weighted GMM
N=200
T=7
Weighted GMM
N=20
T=10
Weighted GMM
N=50
T=10
Weighted GMM
N=100
T=10
Weighted GMM
N=20
T=15
Weighted GMM
N=50
T=15

• References

AHN S.C., SCHMIDT P. (1995). – “Efficient Estimation of Models for Dynamic Panel Data”, *Journal of Econometrics*, 68, pp. 5-27.

AHN S.C., SCHMIDT P. (1997). – “Efficient Estimation of Dynamic Panel Data Models”, *Journal of Econometrics*, 76, pp. 309-321.

ALTONI J.G., SEGAL L.M. (1996). – “Small-Sample Bias in GMM Estimation of Covariance Structures”, *Journal of Business & Economic Statistics*, 14, pp. 353-366.

ANDERSON T.W., HSIAO C. (1981). – “Estimation of Dynamic Models with Error Components”, *Journal of the American Statistical Association*, 77, pp. 598-606.

ANDERSON T.W., HSIAO C. (1982). – “Formulation and Estimation of Dynamic Models Using Panel Data”, *Journal of Econometrics*, 18, pp. 47-82.

ARELLANO M., BOND S. (1991). – “Some Tests of Specification for Panel Data: Monte Carlo Evidence and an Application to Employment Equations”, *Review of Economic Studies*, 58, pp. 277-297.

BEKKER P.A. (1994). – “Alternative Approximations to the Distributions of Instrumental Variable Estimators”, *Econometrica*, 62, pp. 657-681.

CHAMBERLAIN G. (1980). – “Analysis of Covariance with Qualitative Data”, *Review of Economic Studies*, 47, pp. 225-238.

MAGNUS J.R., NEUDECKER H. (1988). – *Matrix Differential Calculus*, Ed. Wiley, Chichester.

NELSON C.R., STARTZ R. (1990a). – “The Distribution of the Instrumental Variables and its *t*-ratio when the Instrument is a Poor One”, *Journal of Business*, 63, pp. 125-140.

NELSON C.R., STARTZ R. (1990b). – “Some Further Results on the Exact Small Sample Properties of the Instrumental Variables Estimator”, *Econometrica*, 58, pp. 967-976.

NEWEY W.K. (1985). – “Generalized Method of Moments Specification Testing”, *Journal of Econometrics*, 28, pp. 229-256.

NICKELL S.J. (1981). – “Biases in Dynamic Models with Fixed Effects”, *Econometrica*, 49, pp. 1417-1426.

SOLOW R.M. (1956). – “A Contribution to the Theory of Economic Growth”, *Quarterly Journal of Economics*, 70, pp. 65-94.

WANSBEEK T.J. (1989). – “Permutation Matrix – II”, *Encyclopedia of Statistical Sciences*, volume 5, S. Kotz and N.L. Johnson Eds, Wiley, New York.

WANSBEEK T.J., BEKKER P.A. (1996). – “On IV, GMM and ML in a Dynamic Panel Data Model”, *Economics Letters*, volume 51, pp. 145-152.

WANSBEEK T.J., KNAAP T. (1998). – “Some Results in Dynamic Fixed Effects Models”, *Mimeo*, unpublished manuscript, SOM, Groningen University.