Working Paper 96-45
Statistics and Econometrics Series 16
July, 1996

Departamento de Economía
Universidad Carlos III de Madrid
Calle Madrid, 126
28903 Getafe (Spain)
Fax (341) 624-98-75

SYMMETRICALLY NORMALIZED INSTRUMENTAL-VARIABLE ESTIMATION USING PANEL DATA

César Alonso-Borrego* and Manuel Arellano **

Abstract

In this paper we discuss the estimation of panel data models with sequential moment restrictions using symmetrically normalized GMM estimators. These estimators are asymptotically equivalent to standard GMM but are invariant to normalization and tend to have a smaller finite sample bias. They also have a very different behaviour compared to standard GMM when the instruments are poor. We study the properties of SN-GMM estimators in relation to GMM, minimum distance and pseudo maximum likelihood estimators for various versions of the AR(1) model with individual effects by mean of simulations. The emphasis is not in assessing the value of enforcing particular restrictions in the model; rather, we wish to evaluate the effects in small samples of using alternative estimating criteria that produce asymptotically equivalent estimators for fixed T and large N. Finally, as an empirical illustration, we estimate by SN-GMM employment and wage equations using panels of UK and Spanish firms.

Keywords: Panel data, instrumental variables, symmetric normalization, autoregressive models, employment equations.

* Departamento de Economía, Departamento de Estadística y Econometría de la Universidad Carlos III de Madrid. ** CEMFI, Madrid

We thank Richard Blundell, Gary Chamberlain, Guido Imbens, Whitney Newey, Enrique Sentana, Jim Stock an seminar audiences at Harvard, Princeton and Northwestern for useful comments. An earlier version of this paper was presented at the ESRC Econometric Study Group Annual Conference, Bristol, July 1994, and at the Econometric Society European Meeting in Maastricht, August 1994.

1. Introduction

In this paper we present instrumental variable estimators of panel data models with predetermined variables subject to a symmetric normalization rule of the coefficients of the endogenous variables. We also evaluate the performance of these techniques for first-order autoregressive models with individual effects by mean of simulations. Lastly, an empirical illustration is provided.

This work is motivated by a concern with the biases of ordinary IV estimators when the instruments are poor. A linear panel data model with predetermined variables, typically estimated by IV techniques, takes the form

$$E(\Delta y_{it} - \Delta x'_{it}\delta \mid z_{i1} \; .. \; z_{it}) = 0, \quad (t=1,..,T; \; i=1,..,N).$$

This formulation includes vector autoregressions and linear Euler equations. The specification of the equation error in first-differences reflects the fact that the analysis is conditional on an unobservable individual effect. Since the number of instruments increases with T, the model generates many overidentifying restrictions even for moderate values of T. However, often the quality of the instruments is poor given that it is usually difficult to predict variables in first differences on the basis of past values of other variables.

The weaker the correlation of the instruments with the endogenous variables, the smaller the amount of information on the structural parameters for a given sample size. However, as it is well documented in the literature on the finite sample properties of simultaneous equations estimators, the way in which this situation is reflected in the distributions of 2SLS and LIML differs substantially, despite the fact that both estimators have the same asymptotic distribution. While the distribution of LIML is centred at the parameter value, 2SLS is biased towards OLS, and in the completely unidentified case converges to a random variable with the OLS probability limit as its central value. On the other hand, LIML has no finite moments regardless of the sample size, and as a consequence its distribution has thicker tails than that of 2SLS and a higher probability of extreme values (see Phillips (1983) for a good survey of the literature). As a result of numerical comparisons of the two distributions involving median-bias, interquartile ranges and rates of approach to normality, Anderson, Kunitomo and Sawa (1982) conclude that LIML is to be strongly preferred to 2SLS, particularly if the number of outside instruments is large. Similar conclusions emerge from the results of asymptotic approximations based on an increasing number of instruments as the sample size tends to infinity; under these sequences, LIML is a consistent estimator but 2SLS is inconsistent (cf. Kunitomo (1980), Morimune (1983) and, more recently, Bekker (1994)).[1] (In our context, these approximations would amount to allowing T to increase to infinity at a chosen rate as opposed to the standard fixed T, large N asymptotics.)

Despite this favourable evidence, LIML has not been used as much in applications as instrumental variables estimators. In the past, LIML was at a disadvantage relative to 2SLS on computational grounds. More fundamentally, applied econometricians have often regarded 2SLS as a more "flexible" choice than LIML from the point of view of the restrictions they were willing to impose on their models. In effect, the IV techniques used for a panel data model with predetermined instruments are not standard 2SLS estimators, since the model gives rise to a system of equations (one for each time period) with a different number of instruments available for each equation. Moreover, concern with heteroskedasticity has lead to consider alternative GMM estimators that use as weighting matrix more robust estimators of the variances and covariances of the orthogonality conditions (following the work of Chamberlain (1982), Hansen (1982) and White (1982)).

In a recent paper, Hillier (1990) shows that the alternative normalization rules adopted by LIML and 2SLS are at the root of their different sampling behaviour. Indeed, Hillier shows that the symmetrically normalized 2SLS estimator (SN-2SLS) has essentially similar properties to those of the LIML estimator. This result, which motivates our focus on symmetrically normalized estimation, is interesting because SN-2SLS, unlike LIML, is a GMM estimator based on structural form orthogonality conditions and therefore it can be readily extended to the nonstandard IV situations that are of interest in panel data models with predetermined variables, while relying on standard GMM asymptotic theory.

To illustrate the situation, let us consider a simple structural equation with a single endogenous explanatory variable and a matrix of instruments Z:

$$y = \beta x + u \tag{1.1}$$

Letting $\hat{y}$ and $\hat{x}$ be the OLS fitted values from the reduced form equations

$$y = Z\pi + v_1$$
$$x = Z\gamma + v_2 \qquad (1.2)$$

the 2SLS estimator of β is given by

$$\hat{\beta}_{2SLS} = \frac{Cov(\hat{x},\hat{y})}{Var(\hat{x})} \equiv \frac{Cov(\hat{x},y)}{Cov(\hat{x},x)}$$

which is not invariant to normalization except in the just-identified case. That is, it differs from the indirect 2SLS estimator:

$$\hat{\beta}_{I2SLS} = \frac{Var(\hat{y})}{Cov(\hat{y},\hat{x})} \equiv \frac{Cov(\hat{y},y)}{Cov(\hat{y},x)}$$

On the other hand, the SN-2SLS estimator is given by the orthogonal regression of $\hat{y}$ on $\hat{x}$, which is invariant to normalization:

$$\hat{\beta}_{SN} = \frac{Cov(\hat{x},\hat{y})}{Var(\hat{x})-\hat{\lambda}} \equiv \frac{Var(\hat{y})-\hat{\lambda}}{Cov(\hat{y},\hat{x})}$$

The statistic $\hat{\lambda}$ is the minimum eigenvalue of the covariance matrix of $\hat{y}$ and $\hat{x}$.

The three estimators have the same first-order asymptotic distribution, but satisfy the inequality

$$|\hat{\beta}_{2SLS}| \le |\hat{\beta}_{SN}| \le |\hat{\beta}_{I2SLS}|$$

Moreover, $\hat{\beta}_{SN}$ can be written as

$$\hat{\beta}_{SN} = \frac{\mathrm{Cov}(\hat{x}+\hat{\beta}_{SN}\hat{y},y)}{\mathrm{Cov}(\hat{x}+\hat{\beta}_{SN}\hat{y},x)}$$

Therefore, 2SLS, I2SLS and SN can all be interpreted as simple IV estimators that use as instruments $\hat{x},\hat{y}$ and $\hat{x} + \hat{\beta}_{SN}\hat{y}$, respectively.

Symmetrically normalized 2SLS can also be given a straightforward interpretation as a GMM or minimum distance estimator, which highlights its relation to LIML. Indeed, both SN-2SLS and LIML are least-squares estimators of the reduced form (1.2) imposing the over-identifying restrictions $\pi=\beta\gamma$. Let us define

$$(\tilde{\beta}_v,\tilde{\gamma}_v) = \underset{\beta,\gamma}{\mathrm{argmin}} \begin{pmatrix} y-Z\gamma\beta \\ x-Z\gamma \end{pmatrix}' (V^{-1}\otimes I) \begin{pmatrix} y-Z\gamma\beta \\ x-Z\gamma \end{pmatrix}$$

$$= \underset{\beta,\gamma}{\mathrm{argmin}} \begin{pmatrix} \hat{\pi}-\gamma\beta \\ \hat{\gamma}-\gamma \end{pmatrix}' (V^{-1}\otimes Z'Z) \begin{pmatrix} \hat{\pi}-\gamma\beta \\ \hat{\gamma}-\gamma \end{pmatrix}$$

Concentrating γ out of the LS criterion we obtain

$$\tilde{\beta}_v = \underset{\beta}{\mathrm{argmin}} \frac{(y-\beta x)'Z(Z'Z)^{-1}Z'(y-\beta x)}{(1,-\beta)'V\begin{pmatrix} 1 \\ -\beta \end{pmatrix}}$$

It turns out that LIML is $\tilde{\beta}_v$ with V equal to the reduced form residual covariance matrix while SN-2SLS is $\tilde{\beta}_v$ with V equal to an identity matrix (cf. Malinvaud (1970), Goldberger and Olkin (1971) and Keller (1975)), so that both LIML and SN-2SLS solve minimum eigenvalue problems. In particular, SN-2SLS is a GMM estimator based on the unit-length orthogonality conditions

$$E\left[\frac{z_i(y_i-\beta x_i)}{(1+\beta^2)^{1/2}}\right] = 0$$

Notice that in spite of V being a matrix scaling factor, the asymptotic distribution of $\hat{\beta}_v$ does not depend on the choice of V. This is so because optimal MD estimators of β based on $(\hat{\pi}-\gamma\beta, \hat{\gamma}-\gamma)$ and on $(\hat{\pi}-\gamma\hat{\beta})$ are asymptotically equivalent, due to the fact that the limiting distribution of optimal MD is invariant to transformations and to the addition of unrestricted moments.

The paper is organized as follows. Section 2 begins with a formulation of the SN-2SLS estimator and its relation to 2SLS and LIML in the general context of a linear structural equation. Next, we present two-step SN-GMM estimators and test statistics of over-identifying restrictions for panel data models with predetermined instruments. Section 3 studies the finite sample properties of SN-GMM estimates in relation to ordinary GMM, minimum distance and pseudo maximum likelihood estimators for various versions of the first-order autoregressive model with individual effects. The objective is not to assess the value of enforcing particular restrictions in the model, but rather to evaluate the effects in small samples, by mean of simulations, of using alternative asymptotically equivalent estimators for fixed T and large N. Section 4 re-estimates the employment equations for a sample of UK firms reported by Arellano and Bond (1991) using symmetrically normalized and indirect GMM estimators. This section further illustrates the techniques by presenting SN-GMM estimates and bootstrap confidence intervals of employment and wage vector autoregressions from a larger panel of Spanish firms. Finally, Section 5 contains the conclusions of the paper.

2. The Symmetrically Normalized Instrumental-Variable Estimator

Preliminaries

We begin this section by providing explicit expressions for 2SLS, LIML and symmetrically normalized 2SLS estimators in order to highlight the algebraic and statistical connections among the three statistics.

Let us consider a standard linear structural equation

$$y_1 = Y_2\beta + Z_1\gamma + u \equiv X\delta + u. \tag{2.1}$$

Also let $Y=(y_1,Y_2)$ be the $n\times(1+p)$ matrix of observations of the endogenous variables, and let $Z=(Z_1,Z_2)$ be the $n\times k$ matrix of instruments, where Z_1 is $n\times k_1$, Z_2 is $n\times k_2$, and $k_2 \geq p$.

The two-stage least squares (2SLS) estimator of δ is given by

$$\hat{\delta}_{2SLS} = \underset{\delta}{\text{argmin}}\ a'W'MWa = (X'MX)^{-1}X'My_1. \tag{2.2}$$

with $W=(Y,Z_1)$, $M=Z(Z'Z)^{-1}Z'$ and $a=(1,-\beta',-\gamma')'$. An expression for the partition of $\hat{\delta}_{2SLS}$ is given by

$$\hat{\beta}_{2SLS} = \underset{\beta}{\text{argmin}}\ b'Y'(M-M_1)Yb = [Y_2'(M-M_1)Y_2]^{-1}Y_2'(M-M_1)y_1$$

$$\hat{\gamma}_{2SLS} = (Z_1'Z_1)^{-1}Z_1'(y_1-Y_2\hat{\beta}_{2SLS})$$

with $b=(1,-\beta')'$ and $M_1=Z_1(Z_1'Z_1)^{-1}Z_1'$.

Similarly, the LIML estimator is given by

$$\hat{\delta}_{LIML} = \underset{\delta}{\text{argmin}}\ \frac{a'W'MWa}{b'\hat{\Omega}b} = [X'(M-\hat{\lambda}(I-M)/n)X]^{-1}X'(M-\hat{\lambda}(I-M)/n)y_1 \qquad (2.3)$$

where $\hat{\lambda}=\min \text{eigen}[Y'(M-M_1)Y\hat{\Omega}^{-1}]$ and $\hat{\Omega}=Y'(I-M)Y/n$, which can be partitioned in accordance with Y as

$$\hat{\Omega} = \begin{bmatrix} \hat{\omega}_{11} & \hat{\omega}_{12} \\ \hat{\omega}_{21} & \hat{\Omega}_{22} \end{bmatrix}.$$

Notice that $\hat{\lambda}\geq 0$. Equally,

$$\hat{\beta}_{LIML} = \underset{\beta}{\text{argmin}}\ \frac{b'Y'(M-M_1)Yb}{b'\hat{\Omega}b} = [Y_2'(M-M_1)Y_2-\hat{\lambda}\hat{\Omega}_{22}]^{-1}[Y_2'(M-M_1)y_1-\hat{\lambda}\hat{\omega}_{21}]$$

$$\hat{\gamma}_{LIML} = (Z_1'Z_1)^{-1}Z_1'(y_1-Y_2\hat{\beta}_{LIML})$$

We define the orthogonal or symmetrically normalized 2SLS estimator (SN-2SLS) to be (see Keller (1975) and Hillier (1990)):

$$\hat{\delta}_{SNM} = \underset{\delta}{\operatorname{argmin}} \frac{a'W'MWa}{b'b} \tag{2.4}$$

Let $Wa^* \equiv Yb^* + Z_1c^* = u^*$ denote equation (2.1) without imposing a normalization rule. With the normalization used by 2SLS $a^*=a$, while with a symmetric normalization of the coefficients of the endogenous variables $a^*=(1+\beta'\beta)^{-1/2}a$. Thus $\hat{\delta}_{SNM}$ is the minimizer of $a^{*\prime}W'MWa^*$ subject to $b^{*\prime}b^*=1$.

Minimizing the criterion (2.4) with respect to γ we obtain a concentrated criterion that only depends on β. This gives us:

$$\hat{\beta}_{SNM} = \underset{\beta}{\operatorname{argmin}} \frac{b'Y'(M-M_1)Yb}{b'b} = [Y_2'(M-M_1)Y_2 - \tilde{\lambda}I]^{-1}Y_2'(M-M_1)y_1$$

$$\hat{\gamma}_{SNM} = (Z_1'Z_1)^{-1}Z_1'(y_1-Y_2\hat{\beta}_{SNM})$$

where $\tilde{\lambda}$=min eigen$[Y'(M-M_1)Y]$. Notice that also $\tilde{\lambda}=\min(a'W'MWa)/b'b$ and that $\tilde{\lambda}\geq 0$. Equivalently,

$$\hat{\delta}_{SNM} = (X'MX - \tilde{\lambda}\Delta)^{-1}X'My_1 \tag{2.5}$$

where $\Delta = \begin{pmatrix} I_p & 0 \\ 0 & 0 \end{pmatrix}$.

In the just identified case, $Z'(y_1-X\hat{\delta}_{2SLS})=0$ which minimizes the three criteria, so that $\hat{\lambda}=\tilde{\lambda}=0$, with the result that 2SLS, LIML and SN-2SLS coincide.

Both $\hat{\delta}_{LIML}$ and $\hat{\delta}_{SNM}$ are invariant to normalization while $\hat{\delta}_{2SLS}$ is not.[2] That is, if the equation is solved for an endogenous variable other than y_1, contrary to the case with 2SLS, the indirect estimates obtained from $\hat{\delta}_{SNM}$ or $\hat{\delta}_{LIML}$ coincide with the direct SNM or LIML estimates, respectively.[3]

The LIML estimator can be regarded as a minimum distance or generalized nonlinear least squares estimator based on the reduced form (see Malinvaud (1970) and Goldberger and Olkin (1971)). Similarly, the SN-2SLS estimator can be viewed as an ordinary nonlinear least squares estimator. To see this, let the reduced form of Y be

$$Y = Z\Pi' + V. \tag{2.6}$$

In view of the partition in Y, the $(1+p)\times k$ matrix of reduced form coefficients can be partitioned as $\Pi'=(\pi_1, \Pi_2')$. In addition, given the structural equation we have

$$\pi_1' = \beta'\Pi_2 + (\gamma',0') \tag{2.7}$$

so that Π is a function of β, γ and Π_2. We can consider NLS estimators of δ and Π_2 that solve

$$(\hat{\delta}_{NLS}, \hat{\Pi}_{2,NLS}) = \text{argmin } tr[V^{-1}(Y-Z\Pi')'(Y-Z\Pi')] \tag{2.8}$$

for particular choices of V. This class of estimators was proposed by Keller (1975). Since Π_2 is not of direct interest we can obtain a concentrated NLS criterion that only depends on δ, which gives $\hat{\delta}_{NLS}$ as the solution to

$$\hat{\delta}_{NLS} = \text{argmin} \frac{a'W'MWa}{b'Vb}. \tag{2.9}$$

Clearly, LIML is $\hat{\delta}_{NLS}$ with $V=\hat{\Omega}$ while SN-2SLS is $\hat{\delta}_{NLS}$ with V=I. The choice of V, provided it is assumed to be bounded in probability or a nonstochastic matrix, leaves the asymptotic distribution of $\hat{\delta}_{NLS}$ unaffected and equal to that of the 2SLS estimator. This result is similar to the one that establishes the equivalence between 2SLS and 3SLS in a system in which there is only one overidentified structural equation.

Symmetrically normalized estimators are attractive alternatives to 2SLS on at least three grounds. Firstly, they tend to have a smaller finite sample bias than the 2SLS estimators. Hillier (1990) shows that for the normal case with p=1 SN-2SLS and LIML are "spherically unbiased" in finite samples.[4] However, 2SLS does not have this property.

Secondly, the concentration of the densities of the symmetrically normalized estimators depends on the quality of the instruments. In the completely unidentified case, as shown by Hillier, these estimators have a uniform distribution on the unit circle. This is in contrast with 2SLS which converges to the same limit as OLS and whose distribution is determined exclusively by the normalization adopted. When the instruments are poor, as well as when the number of instruments is large relative to the sample size, 2SLS tends to provide results that are biased in the direction of OLS and also large discrepancies between "direct" and "indirect" 2SLS when using different normalizations. This situation has been stressed in a number of recent papers (Bekker (1994), Bound, Jaeger and Baker (1995)), Staiger and Stock (1994) and Angrist and Krueger (1995) amongst others). In contrast, with poor instruments the distributions of LIML and SN-2SLS accurately reproduce the fact that the information on the structural parameters is very small.

Thirdly, they are invariant to normalization. SN-2SLS shares these properties in common with LIML; however, one further advantage of SN-2SLS in relation to LIML, is that it is a generalized method of moments estimator based on structural form moment conditions and therefore it can be easily extended to distribution free environments and robust statistics. In particular, it is well suited for application to nonstandard instrumental-variable problems such as those that arise in the context of dynamic and error-in-variables models for panel data.

As the previous discussion reveals, both LIML and SN-2SLS are GMM estimators of δ solved jointly with Π_2 and based on the vector of the reduced form orthogonality conditions:

$$E[z_i \otimes (y_i - \Pi z_i)] = 0 \tag{2.10}$$

where Π is a function of δ and Π_2 (both GMM estimators use a weighting matrix of the form $(V \otimes Z'Z)^{-1}$ with $V=\hat{\Omega}$ for LIML and V=I for SN-2SLS). However, SN-2SLS is also a GMM estimator of δ based on the structural form orthogonality conditions:

$$E\left[\frac{z_i(y_{1i}-x_i'\delta)}{(1+\beta'\beta)^{1/2}}\right] = 0 \tag{2.11}$$

(In the last two expressions, z_i, y_i, y_{1i} and x_i refer to the i-th rows of Z, Y, y_1 and X respectively.)

There is one disadvantage, however, of SN-2SLS relative to the other estimators. In general, the results are not independent of the units in which the variables are measured, so that a sensible choice of the units of scale may be of some importance.[5]

One further useful perspective on SN-2SLS can be obtained by regarding it as a simple IV estimator. The statistic $\hat{\lambda}$ can be written as

$$\hat{\lambda} = y_1'(M-M_1)(y_1-X\hat{\delta}_{SNM})$$

Substituting this expression in the formula for the estimator we obtain

$$\hat{\delta}_{SNM} = (\tilde{Z}'X)^{-1}\tilde{Z}'y_1 \qquad (2.12)$$

where

$$\tilde{Z} = \hat{X} + (M-M_1)y_1\hat{\delta}'_{SNM}\Delta$$

which reduces to $\tilde{Z}=\hat{X}+y_1\hat{\delta}'_{SNM}$ if all the variables in X are endogenous. Remark that for 2SLS we have $\tilde{Z} = \hat{X}$, and more generally for the j-th indirect 2SLS estimator obtained by normalizing to unity the coefficient on the j-th column of Y, we have $\tilde{Z}=\hat{W}(j)$, where $\hat{W}(j)$ coincides with $\hat{W}=(\hat{Y},Z_1)$ except for the j-th column of $\hat{Y}$ which is omitted.

Models for Panel Data

We consider a model with individual effects for panel data given by

$$y_{it} = x'_{it}\delta + u_{it} \qquad (t=1,\dots,T;\ i=1,\dots,N) \qquad (2.13)$$

$$u_{it} = \eta_i + v_{it}$$

The model specifies sequential moment conditions of the form

$$E(v_{it} \mid z_i^t, \eta_i) = 0 \qquad (2.14)$$

were $z_i^t=(z_{i1}' \dots z_{it}')'$ is a vector of instrumental variables.

Thus, this setting is sufficiently general to cover models with strictly exogenous, predetermined and endogenous explanatory variables. We assume that $\{w_i=(y_{i1},\dots,y_{iT},x_{i1}'\dots,x_{iT}',z_{i1}',\dots,z_{iT}')'$, $i=1,\dots N\}$ is a random sample (iid) of size N.

Estimation will be based on a sequence of orthogonality conditions of the form

$$E[z_i^t(y^*_{it} - x^{*\prime}_{it}\delta)] = 0 \qquad (t=1,\dots,T-1) \qquad (2.15)$$

where starred variables denote forward differences or orthogonal deviations of the original variables (e.g. $y^*_{it}=y_{i(t+1)}-y_{it}$).

It is convenient to rewrite the transformed model in the form

$$y_i^* = X_i^*\delta + u_i^*$$

where $y_i^* = (y_{i1}^* \ \dots \ y_{i(T-1)}^*)'$, etc.

The m×1 parameter vector δ is usually estimated by GMM leading to estimators of the form (see Holtz-Eakin, Newey and Rosen (1988), Arellano and Bond (1991), Chamberlain (1992), Arellano and Bover (1995), and Ahn and Schmidt (1995) amongst others):

$$\hat{\delta}_{GMM} = (X^{*\prime} Z\ A_N Z' X^*)^{-1} X^{*\prime} Z\ A_N Z' y^* \qquad (2.16)$$

where $y^* = (y_1^{*\prime} \dots y_N^{*\prime})'$, $X^* = (X_1^{*\prime} \dots X_N^{*\prime})'$ and $Z = (Z_1' \dots Z_N')'$. Z_i is a (T-1)xq block diagonal matrix whose t-th block is z_i^t, and A_N is chosen such that it is a consistent estimate of the inverse of $E(Z_i' u_i^* u_i^{*\prime} Z_i)$. The standard robust choice is

$$A_N = (\textstyle\sum_i Z_i' \tilde{u}_i^* \tilde{u}_i^{*\prime} Z_i)^{-1}$$

where $\tilde{u}_i^*$ is a vector of residuals evaluated using some preliminary consistent estimate of δ. Under very general regularity conditions $\sqrt{N}(\hat{\delta}_{GMM} - \delta)$ is asymptotically normal as $N \to \infty$ and T is fixed, and a consistent estimator of the asymptotic variance of $\hat{\delta}_{GMM}$ is given by

$$\hat{Var}(\hat{\delta}_{GMM}) = (X^{*\prime} Z\ A_N Z' X^*)^{-1} \qquad (2.17)$$

Moreover, the Sargan or GMM statistic of overidentifying restrictions is given by

$$S = \hat{u}^{*\prime} Z\ A_N Z' \hat{u}^* \overset{d}{\to} \chi^2_{q-m}$$

where $\hat{u}^* = y^* - X^*\hat{\delta}_{GMM}$.

Turning to symmetrically normalized GMM (SNM) estimators of δ, let us consider a partition of $X^*=(X_1^*,X_2^*)$ and a corresponding partition of $\delta=(\delta_1',\delta_2')'$ distinguishing between non-exogenous and exogenous variables, such that the m_2 columns of X_2^* are linear combinations of those of Z while the m_1 columns of X_1^* are not.

SNM is the GMM estimator of δ based on the orthogonality conditions

$$E\,\psi(w_i,\delta) = E\left[\frac{Z_i'(y_i^*-X_{1i}^*\delta_1-X_{2i}^*\delta_2)}{(1+\delta_1'\delta_1)^{1/2}}\right] = 0 \tag{2.18}$$

Since $E[\psi(w_i,\delta)\psi'(w_i,\delta)] = E(Z_i'u_i^*u_i^{*\prime}Z_i)/(1+\delta_1'\delta_1)=A_N/(1+\delta_1'\delta_1)$, A_N remains an optimal weighting matrix for the SNM estimator. Therefore,

$$\hat{\delta}_{SNM} = \underset{\delta}{\operatorname{argmin}}\ \frac{(y^*-X^*\delta)'M^*(y^*-X^*\delta)}{(1+\delta_1'\delta_1)} \tag{2.19}$$

where $M^* = ZA_NZ'$. Following our earlier discussion we obtain

$$\hat{\delta}_{1SNM} = \underset{\delta_1}{\operatorname{argmin}}\ \frac{d_1'W_1^{*\prime}(M^*-M_2^*)W_1^*d_1}{d_1'd_1} \tag{2.20}$$

$$\hat{\delta}_{2SNM} = (X_2^{*\prime}M^*X_2^*)^{-1}X_2^{*\prime}M^*(y^*-X_1^*\hat{\delta}_{1SNM}) \tag{2.21}$$

where $W_1^* = (y^*,\ X_1^*)$, $d_1 = (1,-\delta_1')'$ and $M_2^* = M^*X_2^*(X_2^{*\prime}M^*X_2^*)^{-1}X_2^{*\prime}M^*$. So that

$$\delta_{1SNM} = [X_1^{*\prime}(M^*-M_2^*)X_1^* - \tilde{\lambda}I]^{-1} X_1^{*\prime}(M^*-M_2^*)y^* \qquad (2.22)$$

with $\tilde{\lambda}$ = min eigen$[W_1^{*\prime}(M^*-M_2^*)W_1^*]$. A compact expression for $\hat{\delta}_{SNM}$ is given by

$$\hat{\delta}_{SNM} = (X^{*\prime}M^*X^* - \tilde{\lambda}\Delta)^{-1} X^{*\prime}M^*y^* \qquad (2.23)$$

with $\Delta = \begin{pmatrix} I_{m_1} & 0 \\ 0 & 0 \end{pmatrix}$.[6]

Since $\hat{\delta}_{GMM}$ and $\hat{\delta}_{SNM}$ are asymptotically equivalent, $\hat{Var}(\hat{\delta}_{GMM})$ is also a consistent estimate of the asymptotic variance of $\hat{\delta}_{SNM}$. However, an alternative natural estimator of $\hat{Var}(\hat{\delta}_{SNM})$, suggested by the expression above, is

$$\hat{Var}(\hat{\delta}_{SNM}) = (X^{*\prime}M^*X^* - \tilde{\lambda}\Delta)^{-1} \qquad (2.24)$$

Moreover, since $\tilde{\lambda}$ is a minimized optimal GMM criterion it can be used as an alternative test statistic of overidentifying restrictions. We have the result

$$(1 + \hat{\delta}'_{1SNM}\hat{\delta}_{1SNM})\tilde{\lambda} \xrightarrow{d} \chi^2_{q-m} \qquad (2.25)$$

which is asymptotically equivalent to the Sargan test.

The existing evidence from Monte Carlo experiments and empirical analysis point in the direction that, even for moderately large cross-sectional sample sizes, ordinary GMM estimates and their standard errors can be worryingly biased when the instruments are poor. This is typically the case in the context of autoregressive models with individual effects when the roots are close to unity or the contribution of the permanent effect to the total variance is high. If the desirable finite sample properties of symmetrically normalized estimators apply to these environments, $\hat{\delta}_{SNM}$, $\hat{Var}(\hat{\delta}_{SNM})$ and $\tilde{\lambda}$ could provide a useful alternative to estimation and testing.

3. Experimental Comparisons with Alternative Estimators for First Order Autoregressions with Random Effects

The purpose of this section is to study the finite sample properties of the symmetrically normalized GMM estimators in relation to ordinary GMM for various versions of the first-order autoregressive model with individual effects. The IV restrictions implied by these models can also be represented as simple structures on the covariance matrix of the data, and so we can also make comparisons with minimum distance and pseudo maximum likelihood estimators of these covariance structures. The emphasis is not in assessing the value of enforcing particular restrictions in the model, as done for example by Ahn and Schmidt (1995) and Arellano and Bover (1995) for quadratic and stationarity restrictions, respectively. Rather, we wish to evaluate the effects in small samples of using alternative estimating criteria that produce asymptotically equivalent estimators for fixed T and large N. However, since we present results for three different sets of moment restrictions, we shall also be able to make some comparisons across models. We concentrate on a random effects AR(1) model because of its simplicity and the fact that it is a case that has received a great deal of attention in the literature.

Models and Estimators

Let us consider a random sample of individual time-series of size T $y_i^T=(y_{i1},\dots,y_{iT})'$ (i=1,...,N) with second-order moment matrix $E(y_i^T y_i^{T\prime})=\Omega=\{\omega_{ts}\}$. We assume that the joint distribution of y_i^T and the unobservable time-invariant effect η_i satisfies the following assumption:

<u>Assumption A</u>

$$y_{it} = \gamma + \alpha y_{i(t-1)} + \eta_i + v_{it} \qquad (t=2,\dots,T) \tag{3.1}$$

$$E(v_{it}|y_i^{t-1}) = 0 \tag{3.2}$$

where $E(\eta_i)=0$, $E(v_{it}^2)=\sigma_t^2$ and $E(\eta_i^2)=\sigma_\eta^2$.

Notice that since equation (3.1) includes a constant term, it is not restrictive to assume that η_i has zero mean. However, in general $E(\eta_i|y_i^T)$ will be a function of y_i^T. Moreover, the dependence between η_i and v_{it} is not restricted by Assumption A. Another remark is that Assumption A does not rule out the possibility of conditional heteroskedasticity, since $E(v_{it}^2|y_i^{t-1})$ need not coincide with σ_t^2.

Following Arellano and Bond (1991), Assumption A implies (T-2)(T-1)/2 linear moment restrictions of the form

$$E[y_i^{t-2}(\Delta y_{it} - \alpha\Delta y_{i(t-1)})] = 0 \tag{3.3}$$

These restrictions can also be represented as constraints on the elements of Ω. Multiplying (3.1) by y_{is} for $s<t$, and taking expectations gives:

$$\omega_{ts} = \alpha\omega_{(t-1)s} + c_s \qquad (t=2,\ldots T;\ s=1,\ldots,t-1) \tag{3.4}$$

where $c_s = E[y_{is}(\gamma+\eta_i)]$. This means that, given Assumption A, the $T(T+1)/2$ different elements of Ω can be written as functions of the $2T\times1$ parameter vector

$$\theta = (\alpha, c_1, \ldots, c_{T-1}, \omega_{11}, \ldots, \omega_{TT})'$$

We call this moment structure Model 1. Since the moment restrictions in (3.3) are linear in α, they can be used as the basis for a linear GMM estimator of the type discussed in the previous section.

The orthogonality conditions (3.3) are the only restrictions implied by Assumption A on the second-order moments of the data.[7] In particular, with $T=3$ the parameters (α, c_1, c_2) are just-identified as functions of the elements of Ω.

Model 1 is attractive because it is based on minimal assumptions. However, we may be willing to impose additional structure if this conforms to a priori beliefs. One possibility is to assume that the errors v_{it} are mean independent of the individual effect η_i given y_i^{t-1}. This situation gives rise to Assumption B.

Assumption B

$$E(v_{it} | y_i^{t-1}, \eta_i) = 0 \qquad (3.5)$$

Note that Assumption B is more restrictive than Assumption A. When $T \geq 4$, Assumption B implies the following additional T-3 moment restrictions

$$E[(y_{it} - \alpha y_{i(t-1)})(\Delta y_{i(t-1)} - \alpha \Delta y_{i(t-2)})] = 0 \quad (t=4,\ldots,T) \quad (3.6)$$

In effect, we can write

$$E[(y_{it} - \gamma - \alpha y_{i(t-1)} - \eta_i)(\Delta y_{i(t-1)} - \alpha \Delta y_{i(t-2)})] = 0$$

and since $E[(\gamma+\eta_i)\Delta v_{i(t-1)}]=0$ the result follows. GMM estimators of α that exploit these restrictions in addition to those in (3.3) have been considered by Ahn and Schmidt (1995). An alternative representation of the restrictions in (3.6) is in terms of a recursion of the coefficients c_t introduced in (3.5). Multiplying (3.1) by $(\gamma+\eta_i)$ and taking expectations gives:

$$c_t = \alpha c_{t-1} + \phi \qquad (t=2,\ldots,T) \qquad (3.7)$$

where $\phi=\gamma^2+\sigma_\eta^2=E[(\gamma+\eta_i)^2]$, so that $c_1 \ldots c_{T-1}$ can be written in terms of c_1 and ϕ. This gives rise to Model 2 in which Ω depends on the (T+3)x1 parameter vector

$$\theta = (\alpha, \phi, c_1, \omega_{11}, \dots, \omega_{TT})'$$

Notice that with T=3 Assumption B does not imply further restrictions in Ω with the result that α remains just identified relative to the second-order moments.

Other forms of additional structure that can be imposed are various versions of mean or variance stationarity conditions. Assumption C, which requires the change in y_{it} to be mean independent of the individual effect η_i, is a particularly useful mean stationarity condition.

<u>Assumption C</u>

$$E(y_{it} - y_{i(t-1)} | \eta_i) = 0 \qquad (t=2,\dots,T) \tag{3.8}$$

Notice that in combination with Assumption B, Assumption C implies

$$E(y_{it}|\eta_i) = \gamma + \alpha E(y_{it-1}|\eta_i) + \eta_i$$

so that if $E(y_{it}|\eta_i)$ is constant it must be the case that

$$E(y_{it}|\eta_i) = (\gamma+\eta_i)/(1-\alpha) \tag{3.9}$$

and $E(y_{it})=\gamma/(1-\alpha)$.

Relative to Assumption A and Model 1, Assumption C adds the following (T-2) moment restrictions on Ω:

$$E[(y_{it} - \alpha y_{i(t-1)})\Delta y_{i(t-1)}] = 0 \qquad (t=3,\dots,T) \tag{3.10}$$

which were proposed by Arellano and Bover (1995), who developed a linear GMM estimator of α on the basis of (3.3) and (3.10).[8] However, relative to Model 2, Assumption C only adds one moment restriction which can be written as

$$E[(y_{i3} - \alpha y_{i2})\Delta y_{i2}] = 0 \tag{3.11}$$

In terms of the parameters c_t, the implication of Assumption C is that $c_1=\dots=c_{T-1}$ if we move from Model 1, or that $c_1=\phi/(1-\alpha)$ if we move from Model 2. This gives rise to Model 3 in which Ω depends on the $(T+2)\times1$ parameter vector

$$\theta = (\alpha,\phi,\omega_{11},\dots\omega_{TT})'$$

Notice that with T=3, α is overidentified under Assumption C.

The basic specification can be restricted further in various ways. For example, we could consider time series homoskedasticity of the form $E(v_{it}^2)=\sigma^2$ for t=2,...,T and stationarity of the variance of the initial conditions. The combination of these assumptions with Models 2 or 3 would give rise to additional models, some of which have been discussed in detail in the paper by Ahn and Schmidt (1995). However, in the simulations we concentrate in Models 1, 2 and 3 because they embody the restrictions that have been found most useful in applications.

If $E[\psi_j(y_i^T,\alpha)]=0$ denotes the vector of orthogonality conditions available for Model j (j=1,2,3), the symmetrically normalized estimators that we consider are the optimal GMM estimators based on the restrictions $E[\psi_j(y_i,\alpha)/(1+\alpha^2)^{1/2}]=0$. For example, the SNM estimator of α for Model 1 is given by

$$\hat{\alpha}_{SNM,1} = \frac{b_1' A_N b_0}{b_1' A_N b_1 - \tilde{\lambda}} \tag{3.12}$$

where $b_0=N^{-1}\sum_{i=1}^N Z_i'\Delta y_i$, $b_1=N^{-1}\sum_{i=1}^N Z_i'\Delta y_{i(-1)}$, $A_N=(N^{-1}\sum_{i=1}^N Z_i'\Delta\tilde{v}_i\Delta\tilde{v}_i'Z_i)^{-1}$, $\tilde{\lambda}$=min eigen$(B'A_N B)$, $B=(b_0,b_1)$, $\Delta y_i=(\Delta y_{i3}\dots\Delta y_{iT})'$, $\Delta y_{(i-1)}=(\Delta y_{i2}\dots\Delta y_{i(T-1)})'$ and Z_i is a $(T-2)\times(T-2)(T-1)/2$ block diagonal matrix whose sth block is given by y_i^s.

All three models can also be estimated by minimum distance (MD) or by pseudo maximum likelihood (PML) on the basis of the matrix of sample second-order moments $\hat{\Omega}=N^{-1}\sum_{i=1}^N y_i^T y_i^{T\prime}$, and the representations as covariance structures discussed above.

Optimal MD estimators minimize a criterion of the form

$$c_d(\theta) = m(\theta)'V_N^{-1}m(\theta) \tag{3.13}$$

where

$$m(\theta) = \text{vech}[\hat{\Omega} - \Omega(\theta)] = \hat{\omega} - \omega(\theta)$$

and

$$V_N = N^{-1}\sum_{i=1}^N w_i w_i' - \hat{\omega}\hat{\omega}'$$

with $w_i=\text{vech}(y_i^T y_i^{T\prime})$ and $\hat{\omega}=\text{vech}(\hat{\Omega})$.

These estimators have the same asymptotic distribution as the corresponding GMM and SNM estimators. To see this for Model 1, notice that

$$N^{-1}\sum_{i=1}^{N}\psi_1(y_i,\alpha) = b_0 - \alpha b_1 = H_1(\alpha)[\hat{\omega} - \omega(\theta)]$$

where $H_1(\alpha)$ is a $(T-1)(T-2)/2 \times T(T+1)/2$ selection matrix that depends on α. $H_1(\alpha)$ eliminates $(2T-1)$ moments which depend on the 2T parameters contained in θ. Taking into account that the limiting distribution of optimal MD estimators is invariant to transformations and to the addition of unrestricted moments, the asymptotic equivalence between GMM and MD follows.

Turning to PML estimators, one possibility, and the one that we simulate, is to minimize the criterion

$$c_m(\theta) = \log\det\Omega(\theta) + \text{tr}[\Omega^{-1}(\theta)\hat{\Omega}] \tag{3.14}$$

subject to $\Omega(\theta)>0$.[9] The first-order conditions for this PMLE are given by:

$$\left(\frac{\partial\omega(\theta)}{\partial\theta'}\right)' K'[\Omega^{-1}(\theta)\otimes\Omega^{-1}(\theta)]K[\hat{\omega} - \omega(\theta)] = 0$$

where K is a 0-1 matrix such that $K\,\text{vech}(\Omega)=\text{vec}(\Omega)$. It turns out that this PMLE is asymptotically equivalent to the MD estimator that uses $K'(\hat{\Omega}^{-1}\otimes\hat{\Omega}^{-1})K$ as the weighting matrix. Under our Monte Carlo design

$\text{plim}[K'(\hat{\Omega}^{-1}\otimes\hat{\Omega}^{-1})K-V_N^{-1}] = 0$. However, in other environments, such as non-normal or noncentred data, this PMLE would be strictly less efficient asymptotically that the optimal MDE.

An alternative PMLE which is always asymptotically equivalent to the optimal MDE, minimizes

$$c_m^*(\theta) = \log\det(N^{-1}\textstyle\sum_{i=1}^{N}[w_i - \omega(\theta)][w_i - \omega(\theta)]') \qquad (3.15)$$

Since the minimizer of $c_m^*(\theta)$ is equivalent to the iterated MD and it can be expected to be very similar to the MD, it was not included in the simulations.

Monte Carlo Results

We are particularly interested to analyze the behaviour of the estimators in relation with the quality of the instruments. In Model 1 the quality of the instruments basically depends on the values of α and $r=\sigma_\eta^2/\sigma^2$. To illustrate the situation, notice that under stationarity the correlation between Δy_{t-1} and y_{t-2} is given by

$$\rho = -(1 - \alpha)[2(1 - \alpha + (1 + \alpha)r)]^{-1/2}$$

which produces the values

ρ	$\alpha = 0.5$	$\alpha = 0.8$
r = 0	-0.50	-0.32
r = 0.2	-0.39	-0.19
r = 1	-0.25	-0.10

For this reason, we exclude from the simulations models with small values of α, which can be expected to perform relatively well. We consider cases with α=0.5, 0.8, σ^2_η=0, 0.2, 1, T=4, 7 and N=100. The variance of the random error σ^2 is kept equal to unity for all cases. For each experiment we generated 1000 samples of N independent observations of $(y_{i1}, \ldots, y_{iT})$ from the process

$$y_{i1} = (1-\alpha)^{-1}\eta_i + (1-\alpha^2)^{-1/2}v_{i1}$$

$$y_{it} = \alpha y_{i(t-1)} + \eta_i + v_{it} \qquad (t=2,\ldots,T)$$

with $v_i = (v_{i1},\ldots,v_{iT})' \sim N(0,I)$ and $\eta_i \sim N(0,\sigma^2_\eta)$ independent of v_i.

Table 1 reports sample medians, percentage biases, interquartile ranges and median absolute errors for pseudo maximum likelihood (ML), minimum distance (MD), two-step GMM and symmetrically normalized two-step GMM (SNM) estimators for Model 1.[10] The weighting matrices of GMM and SNM are based on optimal one-step GMM residuals as described in Arellano and Bond (1991). In almost every case, SNM is the estimator with the smallest bias and the largest dispersion. When σ^2_η=0 all estimators perform very well, although ML and MD have a smaller interquartile range than GMM and SNM, a difference which is specially noticeable for T=4 (with σ^2_η=0 and α=0.8 the interquartile range of ML or MD is about three times smaller than that of the ordinary or the symmetrically normalized GMM estimators). When σ^2_η=0.2 or 1, the differences in the distributions of GMM and SNM become apparent: the higher σ^2_η or α, the larger the negative bias of GMM for a given T, whereas SNM remains essentially median unbiased. SNM always has a larger interquartile range than GMM, but the differences are small except in the almost unidentified cases (with α=0.8 and T=4). The median absolute errors of GMM and SNM estimates are of a very similar magnitude, although those for GMM tend to be smaller than those for SNM with T=4 and larger with T=7. With T=7, Table 1 clearly indicates that when N=100 there is information in the data to estimate α with sufficient precision but that, contrary to SNM, GMM estimates may still be substantially biased. As far as median bias is concerned, ML and MD are practically unbiased when α=0.5, but exhibit some worryingly large biases when σ^2_η is not zero and α=0.8.

The evidence from Table 1 suggests that Hillier's basic results for ordinary and symmetrically normalized 2SLS estimators may have a wider applicability. In effect, GMM and SNM, unlike 2SLS, are not only functions of the second moments of the data but also of the fourth order moments that enter the weighting matrix of the moment conditions.

Model 1 is the leading case from the point of view that instrumental-variable estimators of structural equations with predetermined instruments tend to rely on orthogonality conditions that are similar to those in Model 1.

Table 2 reports some results for Model 2 that exploits the (T-3) quadratic restrictions given in (3.6) in addition to the linear ones in (3.3). GMM and SNM are asymptotically efficient two-step GMM estimates whose weighting matrix has been calculated using one-step GMM residuals based on the same orthogonality conditions but weighted by an identity matrix. We found that the results are sensitive to the choice of residuals used by the two-step estimates. Unfortunately, in this case, in contrast with the situation for Model 1, there does not seem to be a "natural" choice of one-step GMM estimator that would be asymptotically efficient under classical errors. Another problem is that now GMM is not a linear IV estimator, so that the justification for an estimator based on the downweighted restrictions $E[(1+\alpha^2)^{-1/2}\psi_j(y_i,\alpha)]=0$ becomes dubious. We also tried a version of SNM that only applied the symmetric normalization to the linear orthogonality conditions with very similar results.

In Table 2, ML is, except in two cases, the estimator with the smallest interquartile range and often the one with the smallest bias, with MD trailing ML fairly closely. In drawing comparisons among the estimators, it should be taken into account that the simulated data is normally distributed, so that ML is implicitly using optimally weighted moments with less sampling variability than the methods that rely on higher order moments. On the other hand, ML and MD are subject to the inequality restriction $|\alpha|<1$ while GMM and SNM are not. We experimented with versions of GMM and SNM subject to $|\alpha|<1$ but this did not alter qualitatively the results. Turning to the comparison between GMM and SNM, SNM always has a smaller median bias than GMM, although SNM can also be substantially biased as in the experiment with $\alpha=0.8$, T=7 and $\sigma_\eta^2=1$. Nevertheless, we insist that these results are sensitive to the choice of one-step residuals and further investigation is required.

Table 3 presents the results for Model 3 which makes use of the restrictions derived from Assumptions B and C. This model incorporates the orthogonality conditions from Model 2. However, by adding the stationarity restrictions the entire list of moment conditions admits a linear representation (cf. Ahn and Schmidt (1995)), so that GMM in Table 3 is a linear IV estimator (as proposed by Arellano and Bover (1995)). All the estimators in this Table exhibit small median biases and dispersions, although, as in Table 2, the comparisons favour ML and MD. The differences between GMM and SNM are small in most cases without a clear pattern in the relation, except for the fact that on average SNM estimates are always higher than the GMM estimates.

Both GMM and SNM are two-step estimators based on one-step GMM residuals that use all the orthogonality conditions from Model 3, and the inverse of the second moments of the instruments as the weighting matrix. This one-step estimator is not asymptotically efficient, not even under classical errors. Moreover, the results for GMM and SNM in Table 3 are also sensitive to the choice of one-step residuals. To illustrate the situation, Table 4 reports results for GMM and SNM estimates based on both one-step GMM residuals from Model 1 and one-step residuals from Model 3, but using an identity as the weighting matrix. As an extreme example, the median absolute error of GMM or SNM in Table 3 can be seen to be half of the size of that of GMMb or SNMb in Table A.1 for α=0.8, T=4 and σ^2_η=1. As one would expect, the impact of using Model 1 residuals is more important when Model 1 estimates are highly imprecise. These results suggest that an iterated GMM estimator may often have very different finite sample properties relative to a two-step estimator.

Finally, it is possible to make comparisons across tables. In general, the interquartile ranges become smaller if we move from Table 1 to Table 2 and Table 3. The efficiency gains are particularly important in the cases with α=0.8 and σ^2_η=0.2 or 1. The gains from enforcing stationarity restrictions are always substantial for all the estimators. A puzzling result is that for some experiments the ML and MD estimates of Model 2 have a larger interquartile range than the corresponding estimates for Model 1. However, this result may be related to problems of nonconvergence that we experienced for some of the replications for ML and MD in Model 2.

We have also investigated the finite sample distributions of the standardized GMM and SNM "t-statistics" for Model 1 of the form

$$t_{GMM,1} = \hat{v}_{GMM,1}^{-1/2}(\hat{\alpha}_{GMM,1} - \alpha) \tag{3.16}$$

$$t_{SNM,1} = \hat{v}_{SNM,1}^{-1/2}(\hat{\alpha}_{SNM,1} - \alpha) \tag{3.17}$$

where $\hat{\alpha}_{SNM,1}$ is as defined in (3.12) and $\hat{\alpha}_{GMM,1}$ has a similar expression but with $\tilde{\lambda}$ replaced by zero. The estimated asymptotic variances are given by:

$$\hat{v}_{GMM,1} = 1/(b_1' A_N b_1)$$

$$\hat{v}_{SNM,1} = 1/(b_1' A_N b_1 - \tilde{\lambda})$$

Both $t_{GMM,1}$ and $t_{SNM,1}$ are asymptotically N(0,1). Since the usual tests of hypotheses and confidence intervals rely on this approximation, it is useful to check the accuracy of the approximation for the sample sizes and parameter values considered above.

Table 5 reports finite sample quantiles of the t-statistics based on 10,000 replications. We use a larger number of replications because in this case the 0.9 and 0.95 quantiles in the upper tail of the distribution are of special interest. The median shows that the distributions of the GMM t-statistics are shifted to the left, with the absolute value of the shift increasing with α, σ_{η} and T. In contrast, the distributions of the SNM t-statistics are centered at values very close to zero. Turning to the 0.9 and 0.95 quantiles, when T=4 the differences with the corresponding N(0,1) quantiles are always smaller for the SNM t-statistics than for the GMM, sometimes by a wide margin. When T=7, the normal approximation worsens for both estimators. In that case, however, the upper-tail GMM quantiles tend to be closer to the normal values than those from the SNM t-statistics.

4. Empirical Illustrations

Our first illustration of the previous methods proceeds by re-estimating the employment equations presented by Arellano and Bond (1991) using symmetrically normalized and indirect GMM estimators.

The Arellano-Bond dataset consists on an unbalanced panel of 140 quoted companies from the UK, whose main activity is manufacturing and for which seven, eight or nine continuous annual observations are available for the period 1976-1984.

The models are all log-linear relationships between the number of employees, the average real wage, the stock of capital, a measure of industry output, lagged values of the previous variables, time dummies and company effects. The reader is referred to the Arellano and Bond article for a detailed description of the models and the data.

The first two panels of Table 6 contain the results for two different models estimated in first differences using instrumental variables. Model A includes contemporaneous wage and capital variables, which are treated as endogenous along with the first lag of employment. In this model lagged sales and stocks are used as outside instruments in addition to lags of the endogenous variables included in the equation. Model B only includes lagged values of wages and capital and it could be interpreted as an approximated Euler equation for employment with quadratic adjustment costs. Columns labeled GMM reproduce some of the results obtained by Arellano and Bond. The SNM estimates are calculated as described in Section 2, and for Model A there is an additional column containing indirect GMM estimates that were obtained by normalizing to unity the coefficient of contemporaneous wages. Finally, the third panel of Table 6 presents GMM and SNM estimates of some simple second-order autoregressive models for employment with and without the inclusion of lagged wages.

As Table 6 shows, SNM and indirect GMM estimates are far apart from the direct GMM estimates. These results uncover the fact that the GMM estimates from the dataset of UK firms are probably much less reliable than what their estimated asymptotic standard errors would suggest. Interestingly, the SNM estimates of Model B are more compatible with the Euler equation interpretation than the GMM estimates. For example, in the Euler equation discussed by Arellano and Bond the coefficient on n_{t-1} is given by (2+r) where r is the real discount rate.

Our second empirical illustration is based on a similar but larger balanced panel of 738 Spanish manufacturing companies, for which there are available annual observations for the period 1983-1990 (see the Appendix for a description of these data). We consider a bivariate VAR model for the logarithms of employment and wages. The employment equation contains both lagged employment and lagged wages, while the wage equation only includes its own lags. This model can be regarded as the reduced form of an intertemporal model of employment determination under rational expectations (see Sargent (1978)). To obtain the reduced form, an AR(2) process for log wages is assumed, and the Euler equation in the log of employment for the optimum contingency plans is solved.

Table 7 presents GMM and SNM estimates of the two equations, firstly using only lagged variables in levels as instruments for equations in first-differences (the basic set of moment conditions that we called "Model 1"), and secondly adding lagged variables in first-differences as instruments for equations in levels (that is, including the stationarity restrictions of "Model 3"). For Model 1 we also report estimates of a univariate AR(2) process for employment.

In addition to asymptotic confidence intervals, we calculated 95 percent semiparametric bootstrap confidence intervals based on 1000 replications from the empirical distribution function of the data subject to the moment restrictions (cf. Back and Brown (1993)). Following Brown and Newey (1992) we drew the bootstrap samples from the mass-point distribution that estimated the probability of the i-th observation as

$$\hat{p}_i = 1/(1+\hat{\ell}'\psi(y_i,\hat{\theta}))N$$

where

$$\hat{\ell} = \underset{\ell}{\text{argmin}} \ \frac{1}{N} \Sigma_{i=1}^{N} \log[1+\ell' \psi(y_i,\hat{\theta})]^2$$

and $\psi(y_i,\hat{\theta})$ is the vector of orthogonality conditions for observation i evaluated at the appropriate parameter estimates.

Table 7 contains some interesting results. GMM and SNM estimates of Model 1 are still different from each other but by a smaller margin than the corresponding estimates for the UK panel. The difference becomes even smaller for the univariate employment estimates that are based on half the number of moments used for the estimates in the first two columns. On the other hand, the estimates of Model 3 appear to be more precise, presumably because the additional orthogonality conditions are highly informative. In this case, GMM and SNM estimates provide very similar results. However, the Sargan statistics indicate a clear rejection of the stationarity restrictions in both the employment and the wage equations. It is also noticeable that although bootstrap confidence intervals are always larger than the asymptotic confidence intervals, the differences between the two are generally small.

We re-estimated Model 1 with a random subsample of 200 firms, which is similar to the size of the UK sample. Interestingly, the results (reported in Table 8) are closer to the UK results for similar specifications than those based on the full Spanish sample. In particular, the SNM estimates of the AR(2) model for employment are remarkably stable over the three datasets while standard GMM estimates would be seriously downward biased in the smaller samples. Moreover, the discrepancies between asymptotic and bootstrap confidence intervals in the random subsample were greater than in the full sample.[11]

Finally, we simulated data as close as possible to the AR(2) employment equation, to see if the findings that we obtained with the subsample of 200 companies were substantiated in the Monte Carlo simulations. Random errors and individual effects were generated from independent normal distributions with variances equal to the values estimated from the SNM residuals of the full Spanish sample. Since the estimated time effects showed very little variability, the constant was set to a common value for all periods given by the average estimated time effect in levels, although the estimates in the simulations included time dummies. As a consequence the model was stationary, and we generated (and discarded) 100 preliminary observations for each individual to minimize the impact of initial conditions. The results are reported in Table 9, and confirm the impression conveyed by the real data. The SNM estimates are almost median unbiased, but GMM shows large downward biases, specially when N=200. A comparison in terms of median absolute errors also favours SNM for both sample sizes and parameter estimates. Lastly, looking at the quantiles of the t-ratios shown in the lower panel of Table 9, it appears that the N(0,1) approximation is reasonable for the SNM t-ratios but not for the GMM t-ratios.

5. Conclusions

It has long been established that the lack of finite sample bias is an important advantage of LIML estimators of structural equations over 2SLS, which by contrast have thinner tails than LIML. The bias of 2SLS towards OLS can be specially worrying when the instruments are "poor" and/or the degree of overidentification is large. In practice, this means that while LIML is invariant to normalization, often a 2SLS regression of y on x provides results that are fairly different from those of the (inverted) 2SLS regression of x on y, despite being asymptotically equivalent estimators. However, LIML has not been used much in applications. The reasons for this include a computational disadvantage over 2SLS, concerns with outliers, the fact that 2SLS can be more easily accommodated into the GMM framework, and we suspect that sometimes the use of an implicit prior that favored closeness to OLS when structural coefficients were poorly identified.

There has recently been a renewed interest in the finite sample properties of GMM estimators in various time series and cross-sectional contexts. Several papers have emphasized the role of estimated weighting matrices for the properties of the estimators in small samples, and a number of alternative methods have been considered (eg. Altonji and Segal (1994), Hansen, Heaton and Yaron (1995), Angrist, Imbens and Krueger (1995) or Imbens (1995)). In contrast, in this paper we have focused on the role of normalization rules for the finite sample properties of GMM estimators that make use of standard two-step weighting matrices. Our work is motivated by the results in Hillier (1990), who argued that the alternative normalization rules adopted by LIML and 2SLS are at the basis of their different sampling behaviour. Hillier showed that a symmetrically normalized 2SLS has similar finite sample properties to those of LIML. This result is interesting because, unlike LIML, SN-2SLS is a GMM estimator based on structural form moment conditions and therefore it can be easily extended to distribution free environments and robust statistics.

In particular, SN-2SLS is well suited for application to the nonstandard IV situations that arise in panel data models with predetermined variables, which are the models of interest in this paper. These models are typically estimated in first-differences using all the available lags as instruments. Usually, there is a large number of instruments available, but of poor quality since they tend to be only weakly correlated with the first-differenced endogenous variables that appear in the equation.

In this paper we have presented SN-GMM estimators for dynamic panel data models that are asymptotically equivalent to ordinary optimal GMM estimators. We have also showed how a byproduct of the estimation is a test statistic of overidentifying restrictions, based on a minimum eigenvalue calculation.

We have reported Monte Carlo evidence on the performance of GMM and SN-GMM estimates for a first-order autoregressive model with individual effects. For this model we have considered three alternative sets of moment conditions as discussed by Arellano and Bond (1991), Ahn and Schmidt (1995), and Arellano and Bover (1995). Since for these models, the IV restrictions can be expressed as straightforward structures on the data covariance matrix, using these representations we have also calculated MD and QML estimates for comparisons with the IV estimates. Our findings suggest that Hillier's basic results may have a wider applicability. In most cases, SN-GMM is the estimator with the smallest median bias, and the one with the largest interquartile range. However, the differences in dispersion with ordinary GMM are small except in the almost unidentified cases.

Finally, as an empirical illustration, we have reported estimates of employment and wage equations from UK and Spanish firm panels. The results show that GMM estimates from the (smaller) UK panel can be very unreliable when the degree of overidentification is large. The results from the (larger) Spanish panel produce a closer agreement between ordinary and symmetrically normalized GMM estimates, although there is evidence that there can still be serious biases in GMM estimates. Some of these results are confirmed by simulating data as close as possible to the empirical data. Moment restricted bootstrap confidence intervals show that asymptotic confidence intervals are often over-optimistic, and Sargan tests consistently reject the restrictions implied by the stationarity of initial conditions.

Footnotes

1. Split sample or jackknife IV estimators, however, are also consistent when the number of instruments tends to infinity (cf. Angrist and Krueger (1995) and Angrist, Imbens and Krueger (1995)).

2. Empirical likelihood estimators of the type considered by Qin and Lawless (1994) and Imbens (1995) will also be invariant to normalization due to the invariance property of ML estimators.

3. Notice that if the only explanatory exogenous variable in the equation is a constant term, $\hat{\delta}_{SNM}$ coincides with the orthogonal regression on the fitted values $\hat{Y}$ (cf. Malinvaud (1970) and Anderson (1976)).

4. Meaning that the density of $\hat{\alpha} = \hat{b}/(\hat{b}'\hat{b})^{1/2}$ defined on the unit circle is symmetric about the true points $\pm\alpha=\pm b/(b'b)^{1/2}$ having modes at $\pm\alpha$.

5. This problem does not arise in the autoregressive panel data models discussed below, since in that case the SN-GMM estimator is invariant to units and to normalization.

6. If no columns of X^* are perfectly predictable from Z, or if the entire vector of coefficients is normalized to unity, then $\Delta = I$ and $\tilde{\lambda}=\min \text{eigen}(W^{*\prime}M^*W^*)$, with $W^* = (y^*,X^*)$.

7. However, they are not the only restrictions available since (3.2) also implies that nonlinear functions of y_i^{t-2} are uncorrelated with Δv_{it}. The semiparametric efficiency bound for this model can be obtained from the results in Chamberlain (1992). One reason why estimators based on (3.3) may not be fully efficient asymptotically is that the dependence between η_i and y_i^T may be nonlinear. Another reason would be unaccounted conditional heteroskedasticity.

8. Notice that the (T-2) restrictions in (3.10) can also be written as

$$E[(y_{iT} - \alpha y_{i(T-1)})\Delta y_{i(t-1)}] = 0 \qquad (t=3,\dots,T)$$

For example, we have the identity

$$E(u_{iT}\Delta y_{i(T-2)}) = E[\Delta u_{iT}\Delta y_{i(T-2)}) + E(u_{i(T-1)}\Delta y_{i(T-2)})$$

where $u_{iT}=y_{iT}-\alpha y_{i(T-1)}$.

9. In all cases, optimization with respect to α was conducted over the range $|\alpha|<1$. This was achieved using the reparameterization $\alpha=2\rho/(1+\rho^2)$.

10. Means and standard deviations are not reported since the symmetrically normalized estimators, in common with LIML, can be expected to have infinite moments.

11. Bootstrap standard errors for the UK unbalanced panel were not calculated, since they would depend on a nontrivial specification of the empirical distribution function for the unbalanced observations.

References

Ahn, S. and P. Schmidt (1995), "Efficient Estimation of Models for Dynamic Panel Data, *Journal of Econometrics*, 68, 5-27.

Altonji, J. and L. Segal (1994), "Small-Sample Bias in GMM Estimation of Covariance Structures", NBER Technical Working Paper 156.

Anderson, T.W. (1976), "Estimation of Linear Functional Relationships: Approximate Distributions and Connections with Simultaneous Equations in Econometrics", *Journal of the Royal Statistical Society*, Series B, 38, 1-20.

Anderson, T.W., N. Kunitomo and T. Sawa (1982), "Evaluation of the Distribution Function of the Limited Information Maximum Likelihood Estimator", *Econometrica*, 50, 4, 1009-1027.

Angrist, J.D. and A.B. Krueger (1995), "Split Sample Instrumental Variables Estimates of the Return to Schooling", *Journal of Business and Economic Statistics*, 13, 225-235.

Angrist, J.D., G.W. Imbens and A. Krueger (1995), "Jackknife Instrumental Variables Estimation", NBER Technical Working Paper 172.

Arellano, M. and S.R. Bond (1991), "Some Tests of Specification for Panel Data: Monte Carlo Evidence and an Application to Employment Equations, *Review of Economic Studies*, 58, 277-297.

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

Back, K. and D.P. Brown (1993), "Implied Probabilities in GMM Estimators", *Econometrica*, 61, 971-975.

Bekker, P.A. (1994), "Alternative Approximations to the Distributions of Instrumental Variable Estimators", *Econometrica*, 62, 657-681.

Bound, J., D.A. Jaeger and R. Baker (1995), "The Cure Can Be Worse than the Disease: A Cautionary Tale Regarding Instrumental Variables", *Journal of the American Statistical Association*, April.

Brown, B.W. and W.K. Newey (1992), "Bootstrapping for GMM", mimeo.

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

Chamberlain, G. (1992), "Comment: Sequential Moment Restrictions in Panel Data", *Journal of Business & Economic Statistics*, 10, 20-26.

Goldberger, A.S. and I. Olkin (1971), "A Minimum-Distance Interpretation of Limited-Information Estimation", *Econometrica*, 39, 635-639.

Hansen, L.P. (1982), "Large Sample Properties of Generalized Method of Moments Estimators", *Econometrica*, 50, 1029-1054.

Hansen, L.P., J. Heaton and A. Yaron (1995), "Finite Sample Properties of Some Alternative GMM Estimators", mimeo, Department of Economics, University of Chicago.

Hillier, G.H. (1990), "On the Normalization of Structural Equations: Properties of Direction Estimators", *Econometrica*, 58, 1181-1194.

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

Imbens, G. (1995), "One-step Estimators for Over-identified Generalized Method of Moments Models", mimeo.

Keller, W.J. (1975), "A New Class of Limited-Information Estimators for Simultaneous Equation Systems", *Journal of Econometrics*, 3, 71-92.

Kunitomo, N. (1980), "Asymptotic Expansions of the Distributions of Estimators in a Linear Functional Relationship and Simultaneous Equations", *Journal of the American Statistical Association*, 75, 693-700.

Malinvaud, E. (1970), *Statistical Methods of Econometrics*, North-Holland, Amsterdam.

Morimune, K. (1983), "Approximate Distributions of k-Class Estimators when the Degree of Overidentifiability Is Large Compared with the Sample Size", *Econometrica*, 51, 821-841.

Phillips, P.C.B. (1983), "Exact Small Sample Theory in the Simultaneous Equations Model", in Griliches, Z. and M.D. Intriligator (eds.): *Handbook of Econometrics*, vol. I, North-Holland, Amsterdam, Ch. 8.

Qin, J. and J. Lawless (1994), "Empirical Likelihood and General Estimating Equations", *Annals of Statistics*, 22, 300-325.

Sargent, T.J. (1978), "Estimation of Dynamic Labor Demand Schedules under Rational Expectations", *Journal of Political Economy*, 86.

Staiger, D. and J.H. Stock (1994), "Instrumental Variables Regression with Weak Instruments", NBER Technical Working Paper 151.

White, H. (1982), "Instrumental Variables Regression with Independent Observations", *Econometrica*, 50, 483-499.

Data Appendix

The Spanish dataset is a balanced panel of 738 manufacturing companies recorded in the database of the Bank of Spain's Central Balance Sheet Office from 1983 to 1990. This survey contains information on firm's balance sheets and other complementary information, including data on employment and total wage bill. This survey started in 1982 with the collection of data from large companies with a tendency in subsequent years towards the addition of smaller companies. The database includes both quoted and non quoted firms. The manufacturing firms included in this data set represent more than 40% of the Spanish value added in manufacturing in 1985.

We selected firms reporting information during the whole period 1983-1990 that fulfilled several coherency conditions. All companies with negative values for net worth, capital stock, accumulated depreciation, accounting depreciation, labour costs, employment, sales, output or those whose book value of capital stock jumped by a factor greater than 3 from one year to the next, were dropped from the sample. Finally, we concentrated on non-energy, manufacturing companies with a public share lower than 50 percent.

Variable construction

Employment

Number of employees is dissagregated into permanent employees (those with long-term contracts) and temporary employees (those with short-term contracts). Total employment is calculated as the number of permanent employees, plus the average annual number of temporary employees (number of temporary employees during the year times the average number of weeks worked by temporary employees divided by 52).

Real wage

The measure of the firm's annual average labour costs per employee is computed as the ratio of total wages and salaries (in million Spanish pesetas) to total number of employees. This measure was deflated using Retail Price Indices for each of the subsectors of the manufacturing industry. (Source: Spain's Institute of National Statistics.)

Descriptive statistics

	Mean	Median	Std. deviation	Minimum	Maximum
Employment	310.4	124.0	702.4	10.0	11004.0
Real Wage	1.86	1.75	0.67	0.32	6.66

Table 1

Model 1: Linear restrictions

		α = 0.5				α = 0.8			
		ML	MD	GMM	SNM	ML	MD	GMM	SNM
					T = 4				
$\sigma^2_\eta = 0$									
	median	0.50	0.51	0.49	0.50	0.79	0.80	0.76	0.80
	% bias	0.1	2.1	2.1	0.2	1.5	0.0	5.0	0.3
	iqr	0.11	0.12	0.19	0.19	0.10	0.10	0.28	0.30
	iq80	0.22	0.23	0.36	0.38	0.20	0.21	0.57	0.61
	mae	0.05	0.06	0.09	0.09	0.05	0.05	0.15	0.15
$\sigma^2_\eta = 0.2$									
	median	0.50	0.51	0.47	0.49	0.69	0.71	0.65	0.76
	% bias	0.1	1.3	6.4	1.8	13.7	11.3	18.7	4.5
	iqr	0.19	0.20	0.24	0.25	0.28	0.28	0.47	0.55
	iq80	0.36	0.39	0.47	0.50	0.54	0.58	0.94	1.30
	mae	0.09	0.10	0.12	0.13	0.12	0.11	0.27	0.27
$\sigma^2_\eta = 1$									
	median	0.47	0.49	0.44	0.47	0.65	0.65	0.46	0.65
	% bias	5.5	2.2	12.8	5.3	19.1	19.1	42.6	18.1
	iqr	0.32	0.32	0.35	0.38	0.47	0.48	0.68	0.99
	iq80	0.54	0.56	0.72	0.80	0.90	0.94	1.36	2.59
	mae	0.15	0.16	0.18	0.19	0.18	0.18	0.43	0.51
					T = 7				
$\sigma^2_\eta = 0$									
	median	0.50	0.51	0.48	0.50	0.80	0.81	0.75	0.79
	% bias	0.2	2.0	4.1	0.1	0.5	1.4	5.7	0.8
	iqr	0.08	0.09	0.10	0.10	0.08	0.10	0.13	0.13
	iq80	0.14	0.17	0.19	0.19	0.15	0.17	0.24	0.25
	mae	0.04	0.04	0.05	0.05	0.04	0.05	0.07	0.07
$\sigma^2_\eta = 0.2$									
	median	0.50	0.50	0.47	0.50	0.74	0.74	0.69	0.79
	% bias	0.3	0.1	6.2	0.5	7.7	7.8	13.7	1.7
	iqr	0.10	0.12	0.12	0.12	0.14	0.17	0.20	0.20
	iq80	0.19	0.23	0.23	0.23	0.27	0.34	0.39	0.41
	mae	0.05	0.06	0.06	0.06	0.08	0.09	0.13	0.10
$\sigma^2_\eta = 1$									
	median	0.50	0.50	0.45	0.49	0.72	0.71	0.59	0.77
	% bias	0.6	0.2	9.8	1.4	10.6	11.1	25.9	3.9
	iqr	0.14	0.15	0.14	0.15	0.19	0.22	0.27	0.28
	iq80	0.26	0.29	0.28	0.30	0.37	0.46	0.53	0.59
	mae	0.07	0.08	0.08	0.07	0.10	0.11	0.21	0.15

1,000 replications. N=100, σ_v^2=1.
% bias gives the percentage median bias for all estimates; iqr is the 75th-25th interquartile range; iq80 is the 90th-10th interquantile range; mae denotes the median absolute error.

Table 2

Model 2: Linear and quadratic restrictions

		α = 0.5				α = 0.8			
		ML	MD	GMM	SNM	ML	MD	GMM	SNM
					T = 4				
$\sigma^2_\eta = 0$									
	median	0.50	0.51	0.49	0.50	0.73	0.74	0.75	0.80
	% bias	0.4	1.1	3.0	0.6	8.5	7.2	6.7	0.1
	iqr	0.18	0.19	0.17	0.18	0.19	0.19	0.24	0.27
	iq80	0.33	0.34	0.34	0.36	0.35	0.37	0.50	0.53
	mae	0.09	0.10	0.09	0.09	0.10	0.10	0.13	0.13
$\sigma^2_\eta = 0.2$									
	median	0.49	0.50	0.48	0.51	0.70	0.72	0.71	0.78
	% bias	1.4	0.3	3.3	1.4	12.0	10.3	10.8	2.9
	iqr	0.22	0.22	0.20	0.23	0.22	0.23	0.27	0.33
	iq80	0.39	0.41	0.41	0.46	0.40	0.41	0.56	0.63
	mae	0.11	0.10	0.10	0.11	0.12	0.13	0.16	0.16
$\sigma^2_\eta = 1$									
	median	0.48	0.49	0.48	0.52	0.72	0.73	0.63	0.71
	% bias	4.3	1.7	4.4	3.4	10.3	9.2	21.2	11.2
	iqr	0.23	0.24	0.24	0.27	0.24	0.25	0.33	0.39
	iq80	0.46	0.46	0.49	0.57	0.44	0.45	0.67	0.71
	mae	0.12	0.12	0.12	0.13	0.14	0.13	0.22	0.21
					T = 7				
$\sigma^2_\eta = 0$									
	median	0.50	0.50	0.47	0.49	0.79	0.80	0.74	0.78
	% bias	0.2	1.0	5.0	1.6	1.6	0.1	7.3	2.9
	iqr	0.08	0.10	0.09	0.09	0.11	0.13	0.12	0.13
	iq80	0.16	0.20	0.17	0.18	0.20	0.24	0.23	0.24
	mae	0.04	0.05	0.05	0.05	0.06	0.06	0.08	0.07
$\sigma^2_\eta = 0.2$									
	median	0.50	0.50	0.47	0.49	0.78	0.78	0.68	0.72
	% bias	0.3	0.6	6.5	2.6	2.9	2.4	14.8	10.4
	iqr	0.09	0.11	0.10	0.10	0.11	0.13	0.15	0.16
	iq80	0.16	0.21	0.19	0.20	0.22	0.25	0.32	0.35
	mae	0.04	0.05	0.06	0.05	0.06	0.07	0.13	0.11
$\sigma^2_\eta = 1$									
	median	0.50	0.51	0.45	0.47	0.78	0.78	0.55	0.59
	% bias	0.1	1.4	10.9	6.9	2.8	2.4	30.7	26.8
	iqr	0.08	0.11	0.11	0.11	0.12	0.15	0.23	0.24
	iq80	0.17	0.22	0.22	0.24	0.23	0.26	0.47	0.48
	mae	0.04	0.05	0.07	0.07	0.06	0.07	0.25	0.22

See Notes to Table 1.

Table 3

Model 3: Linear and stationarity restrictions

		α = 0.5				α = 0.8			
		ML	MD	GMM	SNM	ML	MD	GMM	SNM
T = 4									
$\sigma^2_\eta = 0$									
	median	0.50	0.51	0.50	0.51	0.80	0.81	0.79	0.81
	% bias	0.1	1.2	0.8	2.1	0.1	0.7	0.9	1.5
	iqr	0.07	0.07	0.15	0.15	0.05	0.05	0.17	0.17
	iq80	0.12	0.14	0.28	0.28	0.09	0.09	0.32	0.31
	mae	0.03	0.03	0.07	0.07	0.02	0.02	0.08	0.08
$\sigma^2_\eta = 0.2$									
	median	0.50	0.51	0.50	0.51	0.80	0.81	0.79	0.82
	% bias	0.5	1.8	0.9	2.7	0.3	1.3	0.7	2.7
	iqr	0.16	0.19	0.17	0.17	0.19	0.21	0.20	0.19
	iq80	0.30	0.33	0.31	0.32	0.35	0.36	0.37	0.36
	mae	0.08	0.09	0.09	0.09	0.09	0.10	0.10	0.10
$\sigma^2_\eta = 1$									
	median	0.50	0.51	0.52	0.54	0.79	0.82	0.85	0.87
	% bias	0.2	2.3	3.1	8.5	1.3	2.1	5.7	9.2
	iqr	0.20	0.21	0.19	0.20	0.21	0.22	0.19	0.18
	iq80	0.36	0.39	0.36	0.37	0.40	0.40	0.38	0.38
	mae	0.10	0.11	0.09	0.10	0.09	0.10	0.11	0.11
T = 7									
$\sigma^2_\eta = 0$									
	median	0.50	0.51	0.49	0.50	0.80	0.80	0.78	0.80
	% bias	0.1	1.2	2.9	0.1	0.1	0.4	3.0	0.5
	iqr	0.05	0.06	0.08	0.08	0.03	0.04	0.09	0.08
	iq80	0.09	0.11	0.15	0.15	0.06	0.08	0.17	0.16
	mae	0.02	0.03	0.04	0.04	0.02	0.02	0.05	0.04
$\sigma^2_\eta = 0.2$									
	median	0.50	0.50	0.49	0.50	0.80	0.81	0.78	0.80
	% bias	0.3	0.6	2.6	0.9	0.2	1.1	2.4	0.5
	iqr	0.08	0.10	0.09	0.09	0.09	0.12	0.11	0.10
	iq80	0.15	0.20	0.18	0.18	0.17	0.22	0.20	0.19
	mae	0.04	0.05	0.05	0.05	0.05	0.06	0.05	0.05
$\sigma^2_\eta = 1$									
	median	0.50	0.50	0.50	0.51	0.80	0.81	0.83	0.85
	% bias	0.2	0.4	0.7	2.9	0.1	1.8	3.5	5.7
	iqr	0.08	0.11	0.10	0.11	0.11	0.13	0.12	0.11
	iq80	0.16	0.22	0.19	0.20	0.20	0.25	0.22	0.21
	mae	0.04	0.05	0.05	0.05	0.06	0.07	0.07	0.07

See Notes to Table 1.

Table 4

GMM and SNM estimates for Model 3 with alternative residuals

		$\alpha = 0.5$				$\alpha = 0.8$			
		GMMa	SNMa	GMMb	SNMb	GMMa	SNMa	GMMb	SNMb
					T = 4				
$\sigma^2_\eta = 0$									
	median	0.49	0.51	0.49	0.51	0.77	0.81	0.79	0.81
	% bias	2.1	1.1	1.2	1.8	3.2	1.4	1.2	1.4
	iqr	0.16	0.16	0.14	0.15	0.18	0.18	0.18	0.17
	iq80	0.30	0.31	0.28	0.28	0.33	0.34	0.34	0.33
	mae	0.08	0.08	0.07	0.07	0.09	0.09	0.09	0.09
$\sigma^2_\eta = 0.2$									
	median	0.49	0.51	0.49	0.51	0.79	0.83	0.75	0.78
	% bias	1.2	2.6	2.1	1.1	1.3	4.4	6.7	2.1
	iqr	0.18	0.19	0.18	0.18	0.20	0.19	0.26	0.26
	iq80	0.33	0.35	0.32	0.32	0.36	0.36	0.48	0.47
	mae	0.09	0.09	0.09	0.09	0.10	0.11	0.14	0.13
$\sigma^2_\eta = 1$									
	median	0.52	0.55	0.48	0.51	0.86	0.90	0.66	0.73
	% bias	3.3	10.1	4.3	1.8	7.2	12.5	17.3	9.1
	iqr	0.21	0.22	0.21	0.22	0.17	0.16	0.41	0.42
	iq80	0.38	0.39	0.39	0.39	0.34	0.34	0.76	0.85
	mae	0.11	0.12	0.11	0.11	0.10	0.12	0.22	0.21
					T = 7				
$\sigma^2_\eta = 0$									
	median	0.46	0.50	0.49	0.51	0.74	0.81	0.78	0.80
	% bias	7.9	0.4	1.7	1.1	7.9	1.5	2.3	0.0
	iqr	0.10	0.11	0.08	0.08	0.11	0.11	0.09	0.09
	iq80	0.19	0.21	0.15	0.15	0.21	0.21	0.17	0.16
	mae	0.06	0.05	0.04	0.04	0.07	0.06	0.05	0.04
$\sigma^2_\eta = 0.2$									
	median	0.46	0.51	0.49	0.51	0.76	0.84	0.75	0.78
	% bias	8.3	1.5	1.9	1.3	5.5	4.8	6.3	3.1
	iqr	0.11	0.12	0.09	0.09	0.12	0.12	0.13	0.13
	iq80	0.21	0.22	0.17	0.18	0.23	0.23	0.24	0.23
	mae	0.06	0.06	0.05	0.04	0.06	0.07	0.07	0.06
$\sigma^2_\eta = 1$									
	median	0.49	0.54	0.48	0.50	0.83	0.90	0.68	0.70
	% bias	2.7	8.6	3.9	0.3	4.0	12.3	15.5	12.6
	iqr	0.12	0.13	0.11	0.11	0.11	0.10	0.18	0.18
	iq80	0.23	0.25	0.19	0.20	0.22	0.20	0.35	0.34
	mae	0.06	0.07	0.06	0.05	0.06	0.10	0.13	0.11

See Notes to Table 1.
GMMa and SNMa use GMM residuals from Model 3 with weighting identity matrix.
GMMb and SNMb use optimal one-step GMM residuals from Model 1.

Table 5

Model 1: Linear restrictions
Quantiles of the t-statistics

		T = 4				T = 7			
		α = 0.5		α = 0.8		α = 0.5		α = 0.8	
		GMM	SNM	GMM	SNM	GMM	SNM	GMM	SNM
σ^2_η=0									
	0.05	-2.04	-1.94	-2.25	-2.07	-2.49	-2.20	-2.74	-2.18
	0.10	-1.61	-1.51	-1.80	-1.57	-2.01	-1.70	-2.28	-1.74
	0.25	-0.87	-0.77	-1.00	-0.78	-1.22	-0.89	-1.47	-0.92
	0.50	-0.11	0.01	-0.22	0.02	-0.33	0.00	-0.57	-0.03
	0.75	0.58	0.70	0.45	0.69	0.56	0.89	0.28	0.83
	0.90	1.18	1.30	1.00	1.23	1.30	1.64	1.03	1.57
	0.95	1.54	1.65	1.30	1.53	1.76	2.09	1.46	1.99
σ^2_η=0.2									
	0.05	-2.15	-2.04	-2.68	-2.44	-2.62	-2.25	-3.28	-2.34
	0.10	-1.71	-1.58	-2.15	-1.87	-2.11	-1.73	-2.73	-1.83
	0.25	-0.93	-0.81	-1.28	-0.94	-1.30	-0.91	-1.88	-0.98
	0.50	-0.17	-0.02	-0.43	-0.05	-0.41	-0.02	-0.97	-0.11
	0.75	0.54	0.69	0.29	0.68	0.45	0.85	-0.05	0.76
	0.90	1.13	1.28	0.77	1.16	1.24	1.63	0.70	1.50
	0.95	1.44	1.60	0.98	1.42	1.69	2.08	1.13	1.94
σ^2_η=1									
	0.05	-2.36	-2.23	-3.17	-3.09	-2.76	-2.30	-3.82	-2.62
	0.10	-1.83	-1.70	-2.58	-2.46	-2.27	-1.79	-3.26	-2.04
	0.25	-1.09	-0.90	-1.68	-1.42	-1.44	-0.94	-2.35	-1.13
	0.50	-0.25	-0.05	-0.78	-0.28	-0.56	-0.05	-1.37	-0.17
	0.75	0.46	0.68	0.01	0.58	0.32	0.84	-0.43	0.71
	0.90	0.98	1.22	0.50	1.06	1.09	1.59	0.35	1.44
	0.95	1.28	1.50	0.70	1.35	1.51	2.03	0.76	1.86

10,000 replications. N=100, σ_v^2=1.
The 5th, 10th, 25th, 50th, 75th, 90th, and 95th quantiles for the standard normal distribution are, respectively, -1.64, -1.28, -0.67, 0, 0.67, 1.28 and 1.64.

Table 6 (continued)

Employment equations
SNM and GMM estimates from the UK sample

Dependent variable: Δn_{it}	Sample period: 1979-1984 (140 companies)			
	AR(2) Models			
Independent variables	GMM	SNM	GMM	SNM
$\Delta n_{i(t-1)}$	0.691	1.635	0.320	0.827
	(0.051)	(0.074)	(0.053)	(0.065)
$\Delta n_{i(t-2)}$	-0.114	-0.439	0.022	-0.094
	(0.026)	(0.039)	(0.022)	(0.032)
$\Delta w_{i(t-1)}$	0.598	1.958		
	(0.070)	(0.095)		
$\Delta w_{i(t-2)}$	0.013	-0.075		
	(0.036)	(0.053)		
Sargan test (df)	65.9 (50)	71.3 (50)	32.8 (25)	31.3 (25)
R^2 's for IVs:				
$\Delta n_{i(t-1)}$	0.216		0.152	

Notes to Table 6

(i) Time dummies are included in all equations.
(ii) Asymptotic standard errors robust to heteroskedasticity are reported in parentheses.
(iii) All reported estimates are two step.
(iv) Model A treats $\Delta n_{i(t-1)}$, Δw_{it}, $\Delta w_{i(t-1)}$, and Δk_{it} as endogenous. Model B treats $\Delta n_{i(t-1)}$,$\Delta w_{i(t-1)}$, and $\Delta k_{i(t-1)}$ as endogenous.
(v) The instrument set for Models A and B includes lags of employment dated (t-2) and earlier, lags of wages and capital dated (t-2) and (t-3) and the levels and first differences of firm real sales and firm real stocks dated (t-2). The instrument set for all the AR(2) models includes lags of employment dated (t-2) and earlier, and for those in the first two columns also lags of wages dated (t-2) and earlier.
(vi) The R^2 's for the IVs denote the partial R^2 between the instruments and each endogenous explanatory variable once the exogenous variables included in the equation have been partialled out.

<u>Table 7</u>

VAR estimates for employment and wage equations from the Spanish sample

Sample period: 1986-1990 (738 companies)

Independent variables	"Model 1" restrictions			
	GMM	SNM	GMM	SNM
Δn_{it} Equation				
$\Delta n_{i(t-1)}$	0.842 (0.669;1.015) [0.470;1.004]	1.087 (0.894;1.280) [0.729;1.258]	0.748 (0.575;0.921) [0.505;0.989]	0.812 (0.636;0.988) [0.541;0.995]
$\Delta n_{i(t-2)}$	-0.003 (-0.060;0.054) [-0.030;0.137]	-0.074 (-0.140;-0.008) [-0.110;0.067]	0.038 (-0.005;0.081) [-0.012;0.113]	0.030 (-0.015;0.075) [-0.015;0.113]
$\Delta w_{i(t-1)}$	0.078 (-0.086;0.242) [-0.299;0.199]	0.222 (0.046;0.398) [-0.183;0.377]		
$\Delta w_{i(t-2)}$	-0.053 (-0.102;-0.004) [-0.110;0.021]	-0.074 (-0.127;-0.021) [-0.137;-0.003]		
Sargan test (df)	36.9 (36)	37.2 (36)	14.4 (18)	13.5 (18)
<u>R^2's for IVs:</u>				
$\Delta n_{i(t-1)}$	0.033			
$\Delta w_{i(t-1)}$	0.031			
Δw_{it} Equation				
$\Delta w_{i(t-1)}$	0.178 (-0.042;0.398) [-0.170;0.491]	0.228 (-0.008;0.464) [-0.172;0.636]	0.178 (-0.042;0.398) [-0.208;0.542]	0.228 (-0.008;0.464) [-0.237;0.734]
$\Delta w_{i(t-2)}$	-0.012 (-0.081;0.049) [-0.082;0.073]	-0.002 (-0.066;0.062) [-0.076;0.101]	-0.012 (-0.081;0.049) [-0.082;0.082]	-0.002 (-0.066;0.062) [-0.078;0.108]
Sargan test (df)	12.7 (18)	12.9 (18)	12.7 (18)	12.9 (18)
<u>R^2's for IVs:</u>				
$\Delta w_{i(t-1)}$	0.019			

Table 6

Employment equations
SNM and GMM estimates from the UK sample

Dependent variable: Δn_{it} — Sample period: 1979-1984 (140 companies)

Independent variables	Model A GMM	Model A SNM	Model A Indirect GMM[1]	Model B GMM	Model B SNM
$\Delta n_{i(t-1)}$	0.800	1.596	1.214	0.825	2.186
	(0.048)	(0.105)		(0.056)	(0.216)
$\Delta n_{i(t-2)}$	-0.116	-0.384	-0.282	-0.074	-0.455
	(0.021)	(0.045)		(0.020)	(0.077)
Δw_{it}	-0.640	-1.897	-4.638		
	(0.054)	(0.160)			
$\Delta w_{i(t-1)}$	0.564	2.138	1.567	0.431	2.841
	(0.066)	(0.142)		(0.076)	(0.312)
Δk_{it}	0.219	0.238	0.604		
	(0.051)	(0.089)			
$\Delta k_{i(t-1)}$				-0.077	-0.787
				(0.045)	(0.126)
Δys_{it}	0.890	1.747	3.105		
	(0.098)	(0.204)			
$\Delta ys_{i(t-1)}$	-0.874	-2.897	-4.101	-0.115	-2.438
	(0.105)	(0.229)		(0.100)	(0.358)
$\Delta ys_{i(t-2)}$				0.095	1.511
				(0.091)	(0.266)
Sargan test (df)	63.0 (50)	67.1 (50)	62.8 (50)	68.3 (51)	66.5 (51)
R^2 's for IVs:					
$\Delta n_{i(t-1)}$	0.271			0.269	
Δw_{it}	0.193				
$\Delta w_{i(t-1)}$	0.309			0.289	
Δk_{it}	0.108				
$\Delta k_{i(t-1)}$				0.158	

[1]Dependent variable is Δw_{it}.

Table 7 (continued)

VAR estimates for employment and wage equations from the Spanish sample

Sample period: 1986-1990 (738 companies)

Independent variables	"Model 3" restrictions	
	GMM	SNM
Δn_{it} Equation		
$\Delta n_{i(t-1)}$	1.163 (1.112;1.214) [1.064;1.222]	1.208 (1.137;1.279) [1.157;1.370]
$\Delta n_{i(t-2)}$	-0.135 (-0.172;-0.098) [-0.166;-0.044]	-0.142 (-0.185;-0.099) [-0.178;-0.033]
$\Delta w_{i(t-1)}$	0.121 (0.086;0.156) [0.075;0.166]	0.116 (0.077;0.155) [0.054;0.154]
$\Delta w_{i(t-2)}$	-0.132 (-0.171;-0.093) [-0.180;-0.073]	-0.151 (-0.194;-0.108) [-0.232;-0.113]
Sargan test (df)	80.1 (48)	69.1 (48)
Δw_{it} Equation		
$\Delta w_{i(t-1)}$	0.854 (0.815;0.893) [0.790;0.888]	0.873 (0.834;0.912) [0.828;0.926]
$\Delta w_{i(t-2)}$	0.152 (0.105;0.199) [0.107;0.235]	0.138 (0.089;0.187) [0.074;0.207]
Sargan test (df)	71.4 (24)	72.2 (24)

Notes to Table 7

(i) Time dummies are included in all equations.
(ii) All reported estimates are two step.
(iii) The instrument set for all the employment equations under "Model 1" includes lags of employment dated (t-2) and earlier, and for those in the first two columns also lags of wages dated (t-2) and earlier. The instrument set for the wage equation under "Model 1" includes lags of wages dated (t-2) and earlier.
(iv) The R^2 's for the IVs denote the partial R^2 between the instruments and each endogenous explanatory variable once the exogenous variables included in the equation have been partialled out.
(v) 95% asymptotic confidence intervals based on heteroskedasticity-robust standard errors in parentheses; 95% moment-restricted bootstrap confidence intervals in brackets.
(vi) The bootstrap confidence intervals under "Model 1" for the equations in the first two columns are based on a distribution that satisfies a larger set of moment conditions than those in the third and fourth columns. The reason is that the former include lagged wages as instruments for the employment equation, which are absent from the latter.

<u>Table 8</u>

VAR estimates for employment and wage equations from the Spanish sample

Random sample containing 200 companies

Sample period: 1986-1990 (200 companies)

Independent variables	GMM	SNM	GMM	SNM
Δn_{it} Equation				
$\Delta n_{i(t-1)}$	0.788 (0.610;0.966) [0.037;1.234]	1.160 (0.888;1.432) [0.365;1.657]	0.441 (0.167;0.715) [-0.609;0.812]	0.815 (0.509;1.121) [0.237;1.566]
$\Delta n_{i(t-2)}$	-0.042 (-0.109;0.025) [-0.101;0.235]	-0.206 (-0.306;-0.106) [-0.370;0.138]	0.063 (0.002;0.124) [0.000;0.221]	0.003 (-0.062;0.069) [-0.109;0.145]
$\Delta w_{i(t-1)}$	0.337 (0.151;0.523) [-0.238;0.950]	0.650 (0.371;0.929) [0.090;1.759]		
$\Delta w_{i(t-2)}$	0.001 (-0.065;0.067) [-0.098;0.290]	-0.040 (-0.120;0.040) [-0.108;0.254]		
Sargan test (df)	30.2 (36)	23.0 (36)	23.3 (18)	24.3 (18)
<u>R^2's for IVs</u>:				
$\Delta n_{i(t-1)}$	0.064			
$\Delta w_{i(t-1)}$	0.080			
Δw_{it} Equation				
$\Delta w_{i(t-1)}$	-0.612 (-0.984;-0.240) [-3.837;0.314]	-1.198 (-1.442;-0.953) [-4.183;-0.933]	-0.612 (-0.984;-0.240) [-3.766;0.227]	-1.198 (-1.442;-0.953) [-3.989;-0.893]
$\Delta w_{i(t-2)}$	-0.120 (-0.231;-0.009) [-0.715;0.107]	-0.270 (-0.349;-0.191) [-0.878;-0.183]	-0.120 (-0.231;-0.009) [-0.840;0.067]	-0.270 (-0.349;-0.191) [-0.958;-0.160]
Sargan test (df)	17.3 (18)	11.0 (18)	17.3 (18)	11.0 (18)
<u>R^2's for IVs</u>:				
$\Delta w_{i(t-1)}$	0.023			

See Notes to Table 7.

Table 9

Monte Carlo simulations for the AR(2) model for employment
α_1=0.813, α_2 =0.03, γ=0.777, σ^2_η=0.038, σ^2_v=0.01

		N = 738		N = 200	
		GMM	SNM	GMM	SNM
	Summary of estimates				
α_1					
	median	0.72	0.82	0.56	0.82
	% bias	11.6	0.8	30.8	1.1
	iqr	0.15	0.15	0.26	0.27
	iq80	0.28	0.31	0.53	0.58
	mae	0.11	0.08	0.25	0.14
α_2					
	median	0.01	0.03	-0.02	0.02
	% bias	57.7	5.9	165.6	33.8
	iqr	0.03	0.04	0.06	0.08
	iq80	0.07	0.07	0.11	0.14
	mae	0.02	0.02	0.05	0.04
	Quantiles of the t-ratios				
α_1					
	0.10	-2.41	-1.40	-3.43	-1.63
	0.25	-1.75	-0.71	-2.66	-0.76
	0.50	-0.98	0.06	-1.82	0.06
	0.75	-0.20	0.77	-0.96	0.85
	0.90	0.47	1.40	-0.20	1.43
α_2					
	0.10	-2.13	-1.47	-2.85	-1.87
	0.25	-1.37	-0.80	-2.07	-1.13
	0.50	-0.70	-0.07	-1.26	-0.23
	0.75	0.05	0.71	-0.43	0.60
	0.90	0.67	1.30	0.26	1.25

1,000 replications.
% bias gives the percentage median bias for all estimates; iqr is the 75th-25th interquartile range; iq80 is the 90th-10th interquantile range; mae denotes the median absolute error.
The 10th, 25th, 50th, 75th and 90th quantiles for the standard normal distribution are, respectively, -1.28, -0.67, 0, 0.67 and 1.28.