Symmetrically Normalized Instrumental-Variable Estimation Using Panel Data

César Alonso-Borrego
Universidad Carlos III de Madrid,
28903 Getafe, Madrid, Spain

Manuel Arellano
CEMFI, 28014 Madrid, Spain

March 1994
Current version: March 1998

Abstract

We discuss the estimation of linear panel data models with sequential moment restrictions using symmetrically normalized GMM estimators (SNM) and LIML analogues. These estimators are asymptotically equivalent to standard GMM but are invariant to normalization and tend to have a smaller finite sample bias, specially when the instruments are poor. We study their properties in relation to ordinary GMM and minimum distance estimators for AR(1) models with individual effects by mean of simulations. Finally, as empirical illustrations, we estimate by SNM and LIML employment and wage equations using panels of UK and Spanish firms.

Keywords: Dynamic panel data, generalized method of moments, symmetric normalization, autoregressive models, Monte Carlo methods, employment equations.

JEL Classification Code: C23

1 Introduction

This work is motivated by a concern with the finite sample biases of panel data IV estimators when the instruments are poor. A linear panel data model with predetermined variables (like vector autoregressions or linear Euler equations) is typically estimated by IV techniques in first differences using all the available lags of the predetermined variables as instruments. 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 the time series dimension (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) concluded that LIML was to be strongly preferred to 2SLS, particularly if the number of 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). (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 led to consider alternative ("two step") 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 has essentially similar properties to those of the LIML estimator. This result, which motivates our focus on symmetrically normalized estimation, is interesting because the symmetrically normalized 2SLS, unlike LIML, is a GMM estimator based on structural form orthogonality conditions and therefore it can be readily extended to two-step weighting matrices and the nonstandard IV situations that are of interest in dynamic panel data models, while relying on standard GMM asymptotic theory. Specifically, in this paper we discuss both non-robust and robust LIML analogues, and symmetrically normalized GMM estimates in the panel data context.

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

$$y = \beta_o x + u \tag{1}$$

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

$$\begin{aligned} y &= Z\pi_o + v_1 \\ x &= Z\gamma_o + v_2 \end{aligned} \tag{2}$$

and $\widehat{Cov}(\hat{x}, \hat{y}) = \hat{x}'\hat{y}$, etc., the 2SLS estimator of β is given by

$$\hat{\beta}_{2SLS} = \frac{\widehat{Cov}(\hat{x}, \hat{y})}{\widehat{Var}(\hat{x})} \equiv \frac{\widehat{Cov}(\hat{x}, y)}{\widehat{Cov}(\hat{x}, x)} \tag{3}$$

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

$$\hat{\beta}_{I2SLS} = \frac{\widehat{Var}(\hat{y})}{\widehat{Cov}(\hat{y}, \hat{x})} \equiv \frac{\widehat{Cov}(\hat{y}, y)}{\widehat{Cov}(\hat{y}, x)} \tag{4}$$

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

$$\hat{\beta}_{SN} = \frac{\widehat{Cov}(\hat{x}, \hat{y})}{\widehat{Var}(\hat{x}) - \hat{\lambda}} \equiv \frac{\widehat{Var}(\hat{y}) - \hat{\lambda}}{\widehat{Cov}(\hat{y}, \hat{x})} \tag{5}$$

The statistic $\hat{\lambda}$ is the minimum eigenvalue of the sample covariance matrix $(\hat{y} : \hat{x})'(\hat{y} : \hat{x})$ (cf. Malinvaud 1970, and Anderson 1976).

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

$$\left|\hat{\beta}_{2SLS}\right| \leq \left|\hat{\beta}_{SN}\right| \leq \left|\hat{\beta}_{I2SLS}\right| \tag{6}$$

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

$$\hat{\beta}_{SN} = \frac{\widehat{Cov}(\hat{x} + \hat{\beta}_{SN}\hat{y}, y)}{\widehat{Cov}(\hat{x} + \hat{\beta}_{SN}\hat{y}, x)} \tag{7}$$

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 symmetrically normalized 2SLS and LIML are least-squares estimators of the reduced form (2) imposing the over-identifying restrictions $\pi = \beta\gamma$. Let us define

$$(\tilde{\beta}_V, \tilde{\gamma}_V) = \arg\min_{\beta,\gamma} \begin{pmatrix} y - Z\gamma\beta \\ x - Z\gamma \end{pmatrix}' (V^{-1} \otimes I) \begin{pmatrix} y - Z\gamma\beta \\ x - Z\gamma \end{pmatrix} \tag{8}$$

$$= \arg\min_{\beta,\gamma} \begin{pmatrix} \hat{\pi} - \gamma\beta \\ \hat{\gamma} - \gamma \end{pmatrix}' (V^{-1} \otimes Z'Z) \begin{pmatrix} \hat{\pi} - \gamma\beta \\ \hat{\gamma} - \gamma \end{pmatrix} \tag{9}$$

Concentrating γ out of the LS criterion we obtain

$$\tilde{\beta}_V = \arg\min_{\beta} \frac{(y - \beta x)'Z(Z'Z)^{-1}Z'(y - \beta x)}{(1, -\beta')\, V\, (1, -\beta')'} \tag{10}$$

It turns out that LIML is $\tilde{\beta}_V$ with V equal to the reduced form residual covariance matrix while symmetrically normalized 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 symmetrically normalized 2SLS solve minimum eigenvalue problems. In particular, symmetrically normalized 2SLS is a GMM estimator based on the unit-length orthogonality conditions

$$E\left[\frac{z_i(y_i - \beta_o x_i)}{(1 + \beta_o^2)^{1/2}}\right] = 0 \tag{11}$$

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 $(\hat{\pi} - \hat{\gamma\beta})$ 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 one-step symmetrically normalized GMM (SNM) estimator and the non-robust LIML analogue in the context of a linear equation for panel data with sequential moment restrictions. We also present two-step SNM estimators and test statistics of over-identifying restrictions, and compare them with robust LIML analogues. The latter are the "continuously updated GMM" estimators considered by Hansen, Heaton and Yaron (1995). Section 3 studies the finite sample properties of SNM and LIML estimates in relation to ordinary GMM and minimum distance 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, LIML, and indirect GMM estimators. This section further illustrates the techniques by presenting symmetrically normalized 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 Symmetrically Normalized Instrumental Variable Estimation

Let us consider a model with individual effects for panel data given by

$$y_{it} = x_{it}'\delta_o + u_{it} \quad (t = 1, \ldots, T;\ i = 1, \ldots, N) \tag{12}$$

$$u_{it} = \eta_i + v_{it} \tag{13}$$

The model specifies sequential moment conditions of the form

$$E(v_{it}|z_i^t) = 0 \tag{14}$$

where $z_i^t = (z_{i1}' \ldots z_{it}')'$ is a vector of instrumental variables, which may include current or lagged values of y_{it} and x_{it}. Thus, this setting is sufficiently general to cover models with strictly exogenous, predetermined, and endogenous explanatory variables. Observations across individuals are assumed to be independent and identically distributed.

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

$$E\left[z_i^t(y_{it}^* - x_{it}^{*\prime}\delta_o)\right] = 0 \qquad (t = 1, \dots, T-1) \tag{15}$$

where starred variables denote forward orthogonal deviations of the original variables (cf. Arellano and Bover 1995).

It is convenient to rewrite the transformed model in the form

$$y_i^* = X_i^* \delta_o + u_i^* \tag{16}$$

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

The $k \times 1$ parameter vector δ_o 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):

$$\hat{\delta}_{GMM} = (X^{*\prime}ZA_N Z' X^*)^{-1} X^{*\prime} Z A_N Z' y^* \tag{17}$$

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 $(T-1) \times q$ block diagonal matrix whose t-th block is z_i^t, and an optimal choice of A_N is such that it is a consistent estimate of the inverse of $E(Z_i' u_i^* u_i^{*\prime} Z_i)$. Under "classical" errors (that is, when $E(v_{it}^2 \mid z_i^t) = \sigma^2$ and $E(v_{it} v_{i(t+j)} \mid z_i^t) = 0$ for $j > 0$ and all t),

$$E(Z_i' u_i^* u_i^{*\prime} Z_i) = \sigma^2 E(Z_i' Z_i) \tag{18}$$

and hence the "one-step" non-robust choice $A_N = (\hat{\sigma}^2 Z' Z)^{-1}$ is optimal ($\hat{\sigma}^2$, which denotes the residual variance, is irrelevant for estimation but it is kept here for notational convenience). Alternatively, the standard "two-step" robust choice is

$$A_N = \left(\sum_i Z_i' \tilde{u}_i^* \tilde{u}_i^{*\prime} Z_i\right)^{-1} \tag{19}$$

where $\tilde{u}_i^*$ is a vector of residuals evaluated using some preliminary consistent estimate of δ_o. Given identification, $\hat{\delta}_{GMM}$ is consistent and asymptotically normal as $N \to \infty$ for fixed T. In addition, for either choice of A_N, provided the conditions under which they are optimal choices are satisfied, a consistent estimator of the asymptotic variance of $\hat{\delta}_{GMM}$ is given by

$$\widehat{Var}(\hat{\delta}_{GMM}) = (X^{*\prime} Z A_N Z' X^*)^{-1} \tag{20}$$

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

$$S = \hat{u}^{*\prime} Z A_N Z' \hat{u}^* \xrightarrow{d} \chi^2_{q-k} \tag{21}$$

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

Turning to symmetrically normalized GMM (SNM) estimators of δ_o, let us consider a partition of $X^* = (X_1^*, X_2^*)$ and a corresponding partition of $\delta_o = (\delta_{o1}', \delta_{o2}')'$ distinguishing between non-exogenous and exogenous variables, such that the k_2 columns of X_2^* are linear combinations of those of Z while the k_1 columns of X_1^* are not.

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

$$E\psi_i(\delta_o) = E\left[\frac{Z_i'(y_i^* - X_i^* \delta_o)}{(1 + \delta_{o1}' \delta_{o1})^{1/2}}\right] = 0 \tag{22}$$

Since $E[\psi_i(\delta_o)\psi_i'(\delta_o)] = E(Z_i' u_i^* u_i^{*\prime} Z_i)/(1 + \delta_{o1}'\delta_{o1})$, a consistent estimate of the inverse of $E(Z_i' u_i^* u_i^{*\prime} Z_i)$ remains an optimal weighting matrix for the SNM estimator. Therefore,

$$\hat{\delta}_{SNM} = \arg\min_{\delta} \frac{(y^* - X^*\delta)' M (y^* - X^*\delta)}{(1 + \delta_1'\delta_1)} \tag{23}$$

where $M = Z A_N Z'$. Minimizing the criterion with respect to δ_2 we obtain a concentrated criterion that only depends on δ_1. This gives us

$$\begin{aligned} \hat{\delta}_{1SNM} &= \arg\min_{\delta_1} \frac{d_1' W_1^{*\prime}(M - M_2) W_1^* d_1}{d_1' d_1} \\ &= [X_1^{*\prime}(M - M_2) X_1^* - \tilde{\lambda} I]^{-1} X_1^{*\prime}(M - M_2) y^* \end{aligned} \tag{24}$$

$$\hat{\delta}_{2SNM} = (X_2^{*\prime} M X_2^*)^{-1} X_2^{*\prime} M (y^* - X_1^* \hat{\delta}_{1SNM}) \tag{25}$$

where $W_1^* = (y^*, X_1^*)$, $d_1 = (1, -\delta_1')'$, $M_2 = MX_2^*(X_2^{*\prime}MX_2^*)^{-1}X_2^{*\prime}M$, and $\tilde{\lambda} = \min eigen\ [W_1^{*\prime}(M - M_2)W_1^*]$. Notice that also

$$\tilde{\lambda} = \min(y^* - X^*\delta)'M(y^* - X^*\delta)/(1 + \delta_1'\delta_1). \tag{26}$$

Equivalently,

$$\hat{\delta}_{SNM} = (X^{*\prime}MX^* - \tilde{\lambda}\Delta)^{-1}X^{*\prime}My^* \tag{27}$$

with $\Delta = \begin{pmatrix} I_{k_1} & 0 \\ 0 & 0 \end{pmatrix}$ (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 $\tilde{\lambda} = \min\ eigen\ (W^{*\prime}MW^*)$, with $W^* = (y^*, X^*)$). In the just identified case $\tilde{\lambda} = 0$, with the result that GMM and SNM coincide.

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

$$\widehat{Var}(\hat{\delta}_{SNM}) = (X^{*\prime}MX^* - \tilde{\lambda}\Delta)^{-1} \tag{28}$$

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

$$(1 + \hat{\delta}_{1SNM}'\hat{\delta}_{1SNM})\tilde{\lambda} \xrightarrow{d} \chi^2_{q-k} \tag{29}$$

which is asymptotically equivalent to the Sargan test.

Let us now turn to consider LIML analogues or "continuously updated GMM" estimators in the terminology of Hansen et al. (1996). The non-robust LIML analogue $\hat{\delta}_{LIML1}$ minimizes a criterion of the form

$$\ell(\delta) = (y^* - X^*\delta)'Z\ A_N(\delta)Z'(y^* - X^*\delta) \tag{30}$$

with

$$A_N(\delta) = \frac{(Z'Z)^{-1}}{(y^* - X^*\delta)'(y^* - X^*\delta)} \tag{31}$$

The resulting estimator is

$$\widehat{\delta}_{LIML1} = [X^{*}Z(Z'Z)^{-1}Z'X^{*} - \widehat{\ell}X^{*'}X^{*}]^{-1}[X^{*'}Z(Z'Z)^{-1}Z'y^{*} - \widehat{\ell}X^{*'}y^{*}] \tag{32}$$

where, letting $d = (1, -\delta')'$,

$$\begin{aligned} \widehat{\ell} &= \min \frac{d'W^{*'}Z(Z'Z)^{-1}Z'W^{*}d}{d'W^{*'}W^{*}d} \\ &= \min eigen\left[W^{*'}Z(Z'Z)^{-1}Z'W^{*}(W^{*'}W^{*})^{-1}\right] \end{aligned} \tag{33}$$

On the other hand, the robust LIML analogue $\widehat{\delta}_{LIML2}$ minimizes a criterion of the same form as (30) with

$$A_N(\delta) = \left(\sum_{i=1}^{N} Z_i' u_i^{*}(\delta) u_i^{*}(\delta)' Z_i\right)^{-1} \tag{34}$$

where $u_i^{*}(\delta) = y_i^{*} - X_i^{*}\delta$. Therefore, LIML2, unlike LIML1 or the SNM estimators, does not solve a simple minimum eigenvalue problem, and requires the use of numerical optimization methods.

Both the SNM and the LIML analogues are invariant to normalization while the ordinary GMM estimator is not. That is, if the equation is solved for an endogenous variable other than y_i, contrary to the case with ordinary GMM, the indirect estimates obtained from SNM or LIML analogues coincide with the direct SNM or LIML estimates, respectively. (Notice that empirical likelihood estimators of the type considered by Qin and Lawless (1994) and Imbens (1997) will also be invariant to normalization due to the invariance property of ML estimators.)

Symmetrically normalized estimators are potentially attractive alternatives to ordinary GMM on at least two grounds (aside from the desirability of invariance to normalization in its own right). Firstly, they tend to have a smaller finite sample bias than the GMM estimators. Hillier (1990) shows that for the normal case in a standard linear structural equation with two endogenous variables, symmetrically normalized 2SLS and LIML are "spherically unbiased" in finite samples (meaning that the density of $\widehat{a} = \widehat{d}_1/(\widehat{d}_1'\widehat{d}_1)^{1/2}$ defined on the unit circle is symmetric about the true points $\pm a = \pm d_1/(d_1'd_1)^{1/2}$ having modes at $\pm a$). 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 1997, and Angrist and Krueger 1995, amongst others). In contrast, with poor instruments the distributions of LIML and symmetrically normalized 2SLS accurately reproduce the fact that the information on the structural parameters is very small.

Although the LIML analogues and the SNM estimators are asymptotically equivalent (and in the Hillier setting exhibit similar finite sample properties as well), SNM has some disadvantages relative to the other estimators. The main one is that 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. In contrast, ordinary GMM is invariant to units but not to normalization, and LIML is invariant both to units and to normalization. This problem does not arise in the autoregressive panel data models discussed below, since in that case the SNM estimator is invariant to units and to normalization (just because in the autoregressive case a change in the units of the right-hand side variable, leads trivially to a similar change in the units of the left-hand side variable). Another disadvantage of SNM is that the distinction between exogenous and non-exogenous variables is relevant for the specification of the estimator. This is so because in the case of SNM only the length of the coefficient vector for the non-exogenous variables is normalized to unity, and, contrary to LIML, this differs from normalizing to unity the entire coefficient vector. However, SNM does have a computational advantage over LIML when we consider two-step or robust estimators. Indeed, LIML2 or continuously updated GMM no longer solve a minimum eigenvalue problem, whereas two-step SNM only involves simple calculations that are similar to those performed for two-step ordinary GMM. Of course, SNM is limited to linear models, but in such context it is of interest to see if SNM, which is considerably faster and simpler than LIML2, can provide the benefits of the more complicated estimators, and perhaps avoid problems of non-convergence in the case of LIML2.

It is nevertheless possible to consider modified asymptotically efficient two-step SNM estimators that are also invariant to units and yet can be obtained by solving a minimum eigenvalue problem. One such estimator minimizes a criterion of the form

$$\frac{(y^* - X^*\delta)'M(y^* - X^*\delta)}{(y^* - X^*\delta)'(y^* - X^*\delta)} \tag{35}$$

for a two-step choice of M (notice that with the one-step choice this is just the LIML1 criterion).

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 estimators considered above in relation to ordinary GMM for a first-order autoregressive model with individual effects. The IV restrictions implied by various versions of the model can be represented as simple structures on the covariance matrix of the data, and so we can also make comparisons with minimum distance 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), Arellano and Bover (1995), and Blundell and Bond (1998). 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. 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.

3.1 Models and Estimators

Let us consider a random sample of individual time-series of size T, $y_i^T = (y_{i1}, \dots, y_{iT})'$ $(i = 1, \dots, N)$ 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:

Assumption A

$$y_{it} = \alpha y_{i(t-1)} + \eta_i + v_{it} \qquad (t = 2, ..., T) \tag{36}$$

$$E(v_{it} | y_i^{t-1}) = 0 \tag{37}$$

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

Notice that the dependence between η_i and v_{it} is not restricted by Assumption A, nor it is ruled 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

$$E[y_i^{t-2}(\Delta y_{it} - \alpha \Delta y_{i(t-1)})] = 0 \tag{38}$$

These restrictions can also be represented as constraints on the elements of Ω. Multiplying (36) by y_{is} for $s < t$, and taking expectations gives:

$$\omega_{ts} = \alpha \omega_{(t-1)s} + c_s \qquad (t = 2, \dots T; s = 1, \dots, t-1) \tag{39}$$

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

$$\theta = (\alpha, c_1, ..., c_{T-1}, \omega_{11}, \ldots, \omega_{TT})' \tag{40}$$

We call this moment structure Model 1. Since it is a special case of the model in the previous section, all the estimators discussed in Section 2 can be particularized to the present case. Here, however, we express the IV restrictions using errors in first-differences as opposed to orthogonal deviations to simplify the mapping with covariance structures. Notice that with $T = 3$ the parameters (α, c_1, c_2) are just-identified as functions of the elements of Ω.

The orthogonality conditions (38) are the only restrictions implied by Assumption A on the second-order moments of the data. However, they are not the only restrictions available since (37) 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 (38) 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.

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 A$'$.

Assumption A$'$

$$E(v_{it}|y_i^{t-1}, \eta_i) = 0 \tag{41}$$

Note that Assumption A$'$ is more restrictive than Assumption A. When $T \geq 4$, Assumption A$'$ implies the following additional $T-3$ moment restrictions

$$E[(y_{it} - \alpha y_{i(t-1)})(\Delta y_{i(t-1)} - \alpha \Delta y_{i(t-2)})] = 0 \qquad (t = 4, \ldots, T) \tag{42}$$

In effect, we can write $E[(y_{it} - \alpha y_{i(t-1)} - \eta_i)(\Delta y_{i(t-1)} - \alpha \Delta y_{i(t-2)})] = 0$ and since $E(\eta_i \Delta v_{i(t-1)}) = 0$ the result follows.

GMM estimators of α that exploit these restrictions in addition to those in (38) have been considered by Ahn and Schmidt (1995), but since the additional restrictions are nonlinear we do not simulate them here. An alternative representation of the restrictions in (42) is in terms of a recursion of the coefficients c_t introduced in (39). Multiplying (36) by η_i and taking expectations gives:

$$c_t = \alpha c_{t-1} + \phi \qquad (t = 2, \ldots, T) \tag{43}$$

where $\phi = \gamma^2 + \sigma_\eta^2 = E(\eta_i^2)$, so that $c_1 \ldots c_{T-1}$ can be written in terms of c_1 and ϕ. This gives rise to a covariance structure in which Ω depends on the $(T+3) \times 1$ parameter vector $\theta = (\alpha, \phi, c_1, \omega_{11}, \ldots, \omega_{TT})'$. Notice that with $T = 3$ Assumption A$'$ 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 B, which requires the change in y_{it} to be mean independent of the individual effect η_i, is a particularly useful mean stationarity condition.

Assumption B

$$E(y_{it} - y_{i(t-1)} | \eta_i) = 0 \qquad (t = 2, \ldots, T) \tag{44}$$

Notice that given Assumption A, Assumption B implies that $E(y_{it}) = \gamma/(1-\alpha)$.

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

$$E[(y_{it} - \alpha y_{i(t-1)}) \Delta y_{i(t-1)}] = 0 \qquad (t = 3, \ldots, T) \tag{45}$$

which were proposed by Arellano and Bover (1995), who developed a linear GMM estimator of α on the basis of (38) and (45). However, relative to Assumption A$'$, Assumption B only adds one moment restriction which can be written as $E[(y_{i3} - \alpha y_{i2})\Delta y_{i2}] = 0$.

In terms of the parameters c_t, the implication of Assumption B is that $c_1 = \ldots = c_{T-1}$ if we move from Assumption A, or that $c_1 = \phi/(1-\alpha)$ if we move from Assumption A$'$. This gives rise to Model 2, in which Ω depends on the $(T+2) \times 1$ parameter vector

$$\theta = (\alpha, \phi, \omega_{11}, \ldots, \omega_{TT})' \tag{46}$$

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

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 $t = 2, \ldots, T$ and stationarity of the variance of the initial conditions. The combination of these assumptions with the previous ones would give rise to additional models, some of which have been discussed in detail by Ahn and Schmidt (1995). However, in the simulations we concentrate in Models 1 and 2 because they embody linear IV restrictions that have been found most useful in applications. While for Model 1 we shall simulate the robust and non-robust estimators discussed in Section 2, for Model 2 we shall only report robust estimates; that is, the Arellano and Bover (1995) GMM estimator and its symmetrically normalized and continuously updated counterparts. We do so because the combined set of moments in (38) and (45) lack a sequential structure, with the result that there is no simple optimal one-step estimator under "classical" errors.

The coefficient α together with the other free parameters in the covariance structure representations of the previous models can be jointly estimated by minimum distance (MD) on the basis of the matrix of sample second-order moments $\hat{\Omega} = N^{-1} \sum_{i=1}^{N} y_i^T y_i^{T\prime}$. Such estimates have the same asymptotic distribution as the corresponding GMM estimators, but may be cumbersome in more general conditional models since they need to solve a nonlinear optimization problem over a larger parameter space. It is of some interest, however, to compare their finite sample performance with the SNM and LIML estimates of the random effects AR(1) model.

Optimal MD estimators minimize a criterion of the form

$$c_d(\theta) = [\hat{\omega} - \omega(\theta)]' V_N^{-1} [\hat{\omega} - \omega(\theta)] \tag{47}$$

where

$$V_N = N^{-1} \sum_{i=1}^{N} w_i w_i' - \hat{\omega}\hat{\omega}' \tag{48}$$

$\hat{\omega} = vech\ (\hat{\Omega})$ denotes the $T(T+1)/2$ vector containing the elements in the upper triangle of $\hat{\Omega}$, and similarly $\hat{\omega}(\theta) = vech\ [\hat{\Omega}(\theta)]$ and $w_i = vech\ (y_i^T y_i^{T-1\prime})$.

3.2 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

$$\rho = -(1-\alpha)[2(1-\alpha+(1+\alpha)r)]^{-1/2} \tag{49}$$

which is small when α and r are relatively high. 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 $\alpha = 0.5,\ 0.8$, $\sigma_\eta^2 = 0,\ 0.2,\ 1$, $T = 4, 7$ 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

$$y_{i1} = (1-\alpha)^{-1}\eta_i + (1-\alpha^2)^{-1/2} v_{i1} \tag{50}$$

$$y_{it} = \alpha y_{i(t-1)} + \eta_i + v_{it} \quad (t = 2, \ldots, T) \tag{51}$$

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

Tables 1A and 1B report sample medians, percentage biases, interquantile ranges and median absolute errors (*mae*) for GMM, SNM and LIML estimators for Model 1 (means and standard deviations are not reported since the symmetrically normalized estimators can be expected to have infinite moments). Table 1A contains the results for the non-robust estimators and Table 1B for the robust ones. Table 1B also reports the results for the minimum distance estimator, which is also a robust estimator. However, whereas LIM2 and MDE are one-step estimators, GMM2 and SNM2 are calculated in two steps. The weighting matrices of GMM2 and SNM2 are based on GMM1 residuals. SNM1 and LIM1 always have a smaller bias and a larger dispersion than GMM1. When $\sigma_\eta^2 = 0$ all estimators perform well, but when $\sigma_\eta^2 = 0.2$ or 1, the differences in the distributions of GMM1 and the symmetrically normalized estimators become apparent: the higher σ_η^2 or α, the larger the negative bias of GMM1 for a given T, whereas SNM1 remains essentially median unbiased. The behaviour of LIM1 is similar to that of SNM1, although in some cases it shows somewhat larger biases and dispersion. SNM1 and LIM1 have a larger interquartile range than GMM1, but the differences are small except in the almost unidentified cases (with $\alpha = 0.8$ and $T = 4$). The median absolute errors of the three estimators are of a similar magnitude, although those for GMM1 tend to be smaller than those for SNM1 or LIM1 with $T = 4$ and larger with $T = 7$.

Turning to Table 1B, GMM2 and SNM2 exhibit a very similar behaviour to GMM1 and SNM1, respectively. LIM2, which is the robust continuously updated GMM estimator, is virtually median unbiased in all the experiments, although it tends to have a larger *mae* than SNM2. LIM2 was calculated by numerical optimization and we found some instances of nonconvergence. Out of 1000 replications we found 86 cases of nonconvergence for the experiment with $\alpha = 0.8$, $\sigma_\eta^2 = 1$ and $T = 4$, and 7 cases in each of the experiments with $\alpha = 0.8$, $\sigma_\eta^2 = 0.2$, $T = 4$, and $\alpha = 0.8$, $\sigma_\eta^2 = 1$, $T = 7$.

The MDE estimator has a smaller interquartile range than GMM2, SNM2 or

LIM2, a difference which is specially noticeable for $T = 4$ (with $\sigma_\eta^2 = 0$ and $\alpha = 0.8$ the interquartile range of MDE is about three times smaller that that of the other estimators). As far as median bias is concerned, MDE is practically unbiased when $\alpha = 0.5$, but exhibits some larger biases when σ_η^2 is not zero and $\alpha = 0.8$. However, in common with LIM2, we also found a number of cases of non-convergence for MDE, with all the cases arising almost exclusively in the experiments with $\alpha = 0.8$. Specifically, with $\alpha = 0.8$ and $T = 4$, we encountered 36, 46, and 86 cases of non-convergence for $\sigma_\eta^2 = 0$, 0.2, and 1, respectively, whereas with $T = 7$ the number of cases, given in the same order, were 22, 35, and 118.

With $T = 7$, Tables 1A-1B clearly indicate that when $N = 100$ there is information in the data to estimate α with sufficient precision but that, contrary to SNM or LIML, GMM estimates may still be substantially biased.

The evidence from Tables 1A-1B suggests that Hillier's basic results for ordinary and symmetrically normalized 2SLS estimators may have a wider applicability. In effect, GMM2 and SNM2, 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 presents the results for Model 2 which makes use of the restrictions derived from Assumptions A and B. This model incorporates the quadratic orthogonality conditions given in (42). However, by adding the stationarity restrictions the entire list of moment conditions admits a linear representation (cf. Ahn and Schmidt 1995), so that GMM2 in Table 2 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, when there is a difference in *mae* it favours the MDE. The differences between GMM2, SNM2 and LIM2 are small in most cases without a clear pattern in the relation, except for the fact that LIM2 tended to have a smaller bias and it was the estimator with the highest dispersion in all the experiments.

Both GMM2 and SNM2 are two-step estimators based on one-step GMM residuals that use all the orthogonality conditions from Model 2, and the inverse of the second moments of the instruments as the weighting matrix. Notice that this one-step estimator is not asymptotically efficient, not even under classical errors. From calculations based on alternative residuals (not reported), we found that the results for GMM2 and SNM2 were sensitive to the choice of one-step residuals, an issue which does not arise for LIM2 or MDE, as they are calculated in one-step. (We obtained results for GMM2 and SNM2 estimates based on GMM1 residuals from Model 1, and one-step residuals from Model 2, but using an identity as the weighting matrix. As expected, the impact of using Model 1 residuals was more important when Model 1 estimates were highly imprecise.)

Finally, it is possible to make comparisons across tables. The interquartile ranges become smaller if we move from Tables 1A-1B to Table 2. Indeed, the efficiency gains from enforcing stationarity restrictions are always substantial for all the estimators, but they are particularly important in the cases with $\alpha = 0.8$ and $\sigma_\eta^2 = 0.2$ or 1.

We also investigated the finite sample distributions of the standardized one- and two-step GMM, SNM and LIML "t statistics" for Model 1 of the form

$$t = \hat{v}^{-1/2}(\hat{\alpha} - \alpha) \tag{52}$$

where $\widehat{\alpha}$ is an estimator and $\widehat{v}$ is the corresponding estimated asymptotic variance. The t statistics 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.

Tables 3A and 3B report finite sample quantiles of the t statistics based on 10,000 replications for non-robust and robust estimates, respectively. We use a larger number of replications because in this case the 0.90 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^2 and T. In contrast, the distributions of the SNM and LIML t statistics are centred at values that are most of the time very close to zero. Turning to the 0.90 and 0.95 quantiles, when $T = 4$ the differences with the corresponding $N(0, 1)$ quantiles are always smaller for the SNM and LIML t statistics than for the GMM, sometimes by a wide margin. This is true for both non-robust and robust t ratios, although the latter show higher interquantile ranges. When $T = 7$ the contrast between robust and non-robust t ratios becomes more marked. While the normal approximation works reasonably well for SNM1 and LIM1, the distributions of SNM2 and LIM2 exhibit thick tails. The distributions of the GMM t ratios with $T = 7$ remain skewed, but whereas the 0.95 quantiles are very low for GMM1, those for GMM2 tend to be closer to the normal values than those from SNM2 or LIM2.

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.

Table 4A contains 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 GMM2 reproduce some of the results obtained by Arellano and Bond. The SNM2 and LIM2 estimates are calculated as described in Section 2, and for Model A there is an additional column containing indirect GMM2 estimates that were obtained by normalizing to unity the coefficient of contemporaneous wages. Finally, Table 4B presents GMM2, SNM2, and LIM2 estimates of some simple second-order autoregressive models for employment with and without the inclusion of lagged wages.

As Tables 4A-4B show, SNM2, LIM2 and indirect GMM2 estimates are mostly far apart from the direct GMM2 estimates. These results uncover the fact that the GMM2 estimates from the dataset of UK firms are probably much less reliable than what their estimated asymptotic standard errors would suggest. Interestingly, the SNM2 estimates of Model B are more compatible with the Euler equation interpretation than the GMM2 or the LIM2 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 5 presents GMM2, SNM2, and LIM2 estimates of the two equations, 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 Table 6 contains the estimates that add lagged variables in first-differences as instruments for equations in levels (that is, including the stationarity restrictions of "Model 2"). We also report estimates of a univariate AR(2) process for employment for the two models (non-robust estimates are not reported, but are available upon request).

In addition to asymptotic confidence intervals, for GMM2 and SNM2 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

$$\hat{p}_i = 1/[1 + \hat{\mu}'\psi(y_i, \hat{\delta})]N \tag{53}$$

where

$$\hat{\mu} = \arg\min_{\mu} \frac{1}{N} \sum\nolimits_{i=1}^{N} \log[1 + \mu'\psi(y_i, \hat{\delta})]^2 \tag{54}$$

and $\psi(y_i, \hat{\delta})$ is the vector of orthogonality conditions for observation i evaluated at the appropriate parameter estimates. (We were unable to obtain bootstrap confidence intervals for LIM2 due to computing limitations, as each evaluation of LIM2 required numerical optimization over a larger parameter space including time dummies.)

Table 5 contains some interesting results. GMM2 estimates of Model 1 are still different from SNM2 and LIM2 estimates but by a smaller margin than the corresponding estimates for the UK panel. The differences become even smaller for the univariate employment estimates that are based on half the number of moments used for the estimates in the first three columns. On the other hand, the estimates of Model 2 in Table 6 appear to be more precise, presumably because the additional orthogonality conditions are highly informative. In this case, GMM2 and SNM2 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. As for the LIM2 parameter estimates and Sargan statistics, they are similar to GMM2 and SNM2 for the wage equation, but somewhat different for the employment equation. In particular, the first lagged employment coefficient estimate is higher and the Sargan statistic turns out to be much smaller than those for the other estimators.

We re-estimated Model 1 with a random subsample of 200 firms, which is similar to the size of the UK sample. Interestingly, some of the results (reported in Table 7) are closer to the UK results for similar specifications than those based on the full Spanish sample. In particular, the SNM2 estimates of the AR(2) model for employment are remarkably stable over the three datasets while standard GMM2 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. (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.) In contrast, perhaps as a result of a higher probability of outliers in small samples, the LIM2 estimate of the leading coefficient in the AR(2) model for employment was a very small number in the UK sample, and a very large one in the Spanish subsample of 200 firms, whereas it was similar to SNM2 for the full Spanish sample.

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 SNM2 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 for GMM2 and SNM2 are reported in Table 8, and confirm the impression conveyed by the real data (unfortunately, we were unable to simulate LIM2 due to computing limitations). The SNM2 estimates are almost median unbiased, but GMM2 shows large downward biases, specially when $N = 200$. A comparison in terms of median absolute errors also favours SNM2 for both sample sizes and parameter estimates. Lastly, looking at the quantiles of the t ratios shown in the lower panel of Table 8, 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, and the fact that 2SLS can be more easily accommodated into the GMM framework.

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 (e.g. Altonji and Segal 1996, Hansen et al. 1996, Angrist and Krueger 1995, Angrist, Imbens and Krueger 1995, or Imbens 1997). 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, the symmetrically normalized 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, symmetrically normalized 2SLS is well suited for application to the nonstandard IV situations that arise in linear panel data models with predetermined variables, which are the models of interest in this paper. These models are typically estimated in orthogonal deviations or 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 symmetrically normalized GMM (SNM) estimators for dynamic panel data models that are asymptotically equivalent to ordinary optimal GMM estimators. A byproduct of the estimation is a test statistic of overidentifying restrictions, based on a minimum eigenvalue calculation. We have also discussed the relation between robust and non-robust SNM estimators and the LIML analogues. In our context, a non-robust LIML analogue in orthogonal deviations is algebraically equivalent to an ordinary LIML estimator that solves a minimum eigenvalue problem. The robust LIML analogue, however, is the continuously updated GMM estimator proposed by Hansen et al. (1996), which no longer involves a simple minimum eigenvalue calculation.

We have reported Monte Carlo evidence on the performance of non-robust and robust GMM, SNM, and LIML analogue estimates for a first-order autoregressive model with individual effects. For this model we have considered two alternative sets of moment conditions as discussed by Arellano and Bond (1991), 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 minimum distance estimates for comparisons with the IV estimates. Our findings suggest that Hillier's basic results may have a wider applicability. In most cases, the differences in the behaviour of SNM and LIML were small, and both had a smaller median bias, and a larger interquartile range than GMM. However, the differences in dispersion with ordinary GMM were 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 tend to reject the restrictions implied by the stationarity of initial conditions.

Acknowledgements

We thank Richard Blundell, Stephen Bond, Gary Chamberlain, Guido Imbens, Whitney Newey, Enrique Sentana, Jim Stock, Ruey Tsay, an associate editor, three anonymous referees, and seminar audiences at Harvard, Northwestern, Nuffield College, Princeton, and UCL for valuable advice and helpful 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. The first author acknowledges research funding from a CEMFI Ph.D. scholarship and from the Spanish DGES, Grant PB95-0292.

Appendix: Data Description

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

References

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

Altonji, J., and Segal, L. (1996), "Small-Sample Bias in GMM Estimation of Covariance Structures", *Journal of Business & Economic Statistics*, 14, 353-366.

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

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

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

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

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

Back, K., and Brown, D.P. (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.

Blundell, R., and Bond, S. (1998), "Initial Conditions and Moment Restrictions in Dynamic Panel Data Models", *Journal of Econometrics*, in press.

Bound, J., Jaeger, D.A., and Baker, R. (1995), "Problems with Instrumental Variables Estimation When the Correlation Between the Instruments and the Endogenous Explanatory Variable is Weak", *Journal of the American Statistical Association*, 90, 443-450.

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

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 Olkin, I. (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., Heaton, J., and Yaron, A. (1996), "Finite Sample Properties of Some Alternative GMM Estimators", *Journal of Business & Economic Statistics*, 14, 262-280.

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

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

Imbens, G. (1997), "One-step Estimators for Over-identified Generalized Method of Moments Models", *Review of Economic Studies*, 64, 359-383.

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 Lawless, J. (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, 1009-1044.

Staiger, D., and Stock, J.H. (1997), "Instrumental Variables Regression with Weak Instruments", *Econometrica*, 65, 557-586.

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

Table 1A
Model 1: Non-robust estimates

		$\alpha = 0.5$			$\alpha = 0.8$		
		GMM1	SNM1	LIM1	GMM1	SNM1	LIM1
				$T = 4$			
$\sigma_\eta^2 = 0$							
	median	0.49	0.50	0.50	0.76	0.80	0.80
	% bias	2.5	0.3	0.6	5.6	0.1	0.1
	iqr	0.18	0.19	0.19	0.28	0.29	0.29
	iq80	0.35	0.36	0.36	0.56	0.61	0.61
	mae	0.09	0.09	0.09	0.15	0.15	0.15
$\sigma_\eta^2 = 0.2$							
	median	0.47	0.49	0.49	0.66	0.77	0.77
	% bias	6.9	1.7	1.7	17.8	3.7	4.1
	iqr	0.23	0.25	0.24	0.45	0.57	0.58
	iq80	0.44	0.47	0.47	0.93	1.26	1.29
	mae	0.12	0.12	0.12	0.25	0.28	0.29
$\sigma_\eta^2 = 1$							
	median	0.43	0.48	0.48	0.44	0.65	0.61
	% bias	14.8	3.8	3.1	44.7	19.0	23.8
	iqr	0.33	0.36	0.36	0.67	0.95	1.02
	iq80	0.68	0.77	0.77	1.39	2.81	2.89
	mae	0.18	0.18	0.18	0.44	0.50	0.53
				$T = 7$			
$\sigma_\eta^2 = 0$							
	median	0.47	0.50	0.49	0.75	0.80	0.79
	% bias	5.0	0.7	2.0	6.0	0.3	1.1
	iqr	0.09	0.09	0.09	0.11	0.12	0.12
	iq80	0.16	0.17	0.17	0.22	0.23	0.24
	mae	0.05	0.04	0.04	0.07	0.06	0.06
$\sigma_\eta^2 = 0.2$							
	median	0.47	0.50	0.49	0.70	0.81	0.78
	% bias	6.7	0.8	1.8	13.0	1.2	2.7
	iqr	0.11	0.11	0.11	0.18	0.18	0.21
	iq80	0.20	0.21	0.21	0.34	0.39	0.45
	mae	0.06	0.06	0.06	0.12	0.09	0.11
$\sigma_\eta^2 = 1$							
	median	0.45	0.50	0.48	0.61	0.82	0.74
	% bias	10.4	1.0	3.3	24.0	3.0	8.1
	iqr	0.13	0.14	0.14	0.23	0.26	0.38
	iq80	0.24	0.26	0.27	0.45	0.54	0.86
	mae	0.07	0.07	0.07	0.20	0.13	0.19

Table 1B

Model 1: Robust estimates

	$\alpha = 0.5$				$\alpha = 0.8$			
	GMM2	SNM2	LIM2	MDE	GMM2	SNM2	LIM2	MDE
				$T = 4$				
$\sigma_\eta^2 = 0$								
median	0.49	0.50	0.51	0.51	0.76	0.80	0.81	0.80
% bias	2.1	0.2	1.6	2.1	4.9	0.3	1.7	0.0
iqr	0.19	0.19	0.19	0.12	0.29	0.30	0.31	0.10
iq80	0.36	0.38	0.38	0.23	0.58	0.62	0.63	0.21
mae	0.09	0.09	0.09	0.06	0.15	0.15	0.16	0.05
$\sigma_\eta^2 = 0.2$								
median	0.47	0.49	0.50	0.51	0.65	0.76	0.84	0.71
% bias	6.5	1.8	0.3	1.3	19.0	4.6	5.1	11.3
iqr	0.24	0.25	0.25	0.20	0.47	0.55	0.56	0.28
iq80	0.47	0.50	0.51	0.39	0.97	1.33	1.23	0.58
mae	0.12	0.13	0.13	0.10	0.27	0.28	0.28	0.11
$\sigma_\eta^2 = 1$								
median	0.44	0.47	0.50	0.49	0.45	0.64	0.82	0.65
% bias	12.8	5.4	0.5	2.2	43.6	19.5	2.9	19.1
iqr	0.35	0.38	0.38	0.32	0.70	1.03	0.94	0.48
iq80	0.75	0.80	0.80	0.56	1.53	2.82	2.22	0.94
mae	0.18	0.19	0.19	0.16	0.46	0.54	0.47	0.18
				$T = 7$				
$\sigma_\eta^2 = 0$								
median	0.48	0.50	0.50	0.51	0.75	0.79	0.80	0.81
% bias	4.3	0.4	0.6	2.0	5.7	0.8	0.1	1.4
iqr	0.10	0.10	0.10	0.09	0.13	0.13	0.14	0.10
iq80	0.18	0.19	0.21	0.17	0.24	0.25	0.28	0.17
mae	0.05	0.05	0.05	0.04	0.07	0.07	0.07	0.05
$\sigma_\eta^2 = 0.2$								
median	0.47	0.50	0.50	0.50	0.69	0.79	0.81	0.74
% bias	6.2	0.5	0.4	0.1	13.7	1.7	0.9	7.8
iqr	0.12	0.12	0.13	0.12	0.20	0.20	0.24	0.17
iq80	0.23	0.23	0.26	0.23	0.39	0.41	0.51	0.34
mae	0.06	0.06	0.06	0.06	0.13	0.10	0.12	0.09
$\sigma_\eta^2 = 1$								
median	0.45	0.49	0.50	0.50	0.59	0.77	0.80	0.71
% bias	9.8	1.5	0.0	0.2	26.0	3.9	0.1	11.1
iqr	0.14	0.15	0.16	0.15	0.27	0.28	0.36	0.22
iq80	0.28	0.30	0.33	0.29	0.53	0.59	0.80	0.46
mae	0.08	0.07	0.08	0.08	0.22	0.15	0.18	0.11

Notes to Tables 1A and 1B:

1,000 replications. $N = 100$, $\sigma_v^2 = 1$.

% bias gives the percentage median bias for all the 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: Robust estimates

	$\alpha = 0.5$				$\alpha = 0.8$			
	GMM2	SNM2	LIM2	MDE	GMM2	SNM2	LIM2	MDE
				$T = 4$				
$\sigma_\eta^2 = 0$								
median	0.50	0.51	0.51	0.51	0.79	0.81	0.81	0.81
% bias	0.8	2.0	1.9	1.2	0.9	1.5	1.7	0.7
iqr	0.15	0.15	0.15	0.07	0.17	0.17	0.17	0.05
iq80	0.28	0.28	0.29	0.14	0.32	0.31	0.33	0.09
mae	0.07	0.07	0.08	0.03	0.08	0.08	0.09	0.02
$\sigma_\eta^2 = 0.2$								
median	0.50	0.51	0.51	0.51	0.79	0.82	0.81	0.81
% bias	0.9	2.5	1.9	1.8	0.7	2.7	1.5	1.3
iqr	0.17	0.17	0.19	0.19	0.20	0.19	0.22	0.21
iq80	0.31	0.32	0.33	0.33	0.37	0.36	0.40	0.36
mae	0.09	0.09	0.09	0.09	0.10	0.10	0.11	0.10
$\sigma_\eta^2 = 1$								
median	0.52	0.54	0.51	0.51	0.85	0.87	0.81	0.82
% bias	3.1	8.4	1.93	2.3	5.7	9.2	1.0	2.1
iqr	0.19	0.20	0.21	0.21	0.19	0.18	0.25	0.22
iq80	0.36	0.37	0.39	0.39	0.38	0.37	0.43	0.40
mae	0.09	0.10	0.11	0.11	0.11	0.11	0.12	0.10
				$T = 7$				
$\sigma_\eta^2 = 0$								
median	0.49	0.50	0.50	0.51	0.78	0.80	0.80	0.80
% bias	2.9	0.1	0.6	1.2	3.0	0.5	0.6	0.4
iqr	0.08	0.08	0.09	0.06	0.09	0.08	0.09	0.04
iq80	0.15	0.16	0.17	0.11	0.17	0.16	0.18	0.08
mae	0.04	0.04	0.04	0.03	0.05	0.04	0.05	0.02
$\sigma_\eta^2 = 0.2$								
median	0.49	0.50	0.50	0.50	0.78	0.80	0.81	0.81
% bias	2.6	0.9	0.6	0.6	2.4	0.5	1.1	1.1
iqr	0.09	0.09	0.10	0.10	0.11	0.10	0.12	0.12
iq80	0.18	0.18	0.20	0.20	0.20	0.19	0.22	0.22
mae	0.05	0.05	0.05	0.05	0.05	0.05	0.06	0.06
$\sigma_\eta^2 = 1$								
median	0.50	0.51	0.50	0.50	0.83	0.85	0.81	0.81
% bias	0.7	2.9	0.2	0.4	3.5	5.7	0.7	1.8
iqr	0.10	0.11	0.11	0.11	0.12	0.11	0.14	0.13
iq80	0.19	0.20	0.22	0.22	0.22	0.21	0.25	0.25
mae	0.05	0.05	0.05	0.05	0.07	0.07	0.07	0.07

See Notes to Tables 1A and 1B.

Table 3A
Model 1: Non-robust estimates
Quantiles of the t statistics

	$T=4$						$T=7$					
	$\alpha=0.5$			$\alpha=0.8$			$\alpha=0.5$			$\alpha=0.8$		
	GMM1	SNM1	LIM1	GMM1	SNM1	LIM1	GMM1	SNM1	LIM1	GMM1	SNM1	LIM1
$\sigma_\eta^2=0$												
0.05	-1.97	-1.84	-1.87	-2.16	-1.90	-2.03	-2.04	-1.66	-1.84	-2.26	-1.62	-1.95
0.10	-1.54	-1.42	-1.44	-1.74	-1.46	-1.56	-1.65	-1.27	-1.46	-1.87	-1.25	-1.51
0.25	-0.86	-0.74	-0.75	-0.98	-0.73	-0.78	-1.01	-0.64	-0.79	-1.23	-0.61	-0.82
0.50	-0.13	-0.02	-0.01	-0.25	0.00	-0.01	-0.32	0.04	-0.08	-0.53	0.06	-0.07
0.75	0.53	0.62	0.64	0.41	0.59	0.64	0.37	0.74	0.64	0.17	0.73	0.68
0.90	1.08	1.17	1.22	0.93	1.06	1.15	0.98	1.33	1.26	0.75	1.29	1.30
0.95	1.41	1.48	1.53	1.20	1.30	1.41	1.33	1.70	1.65	1.10	1.61	1.67
$\sigma_\eta^2=0.2$												
0.05	-2.05	-1.89	-1.95	-2.39	-2.00	-2.38	-2.12	-1.65	-1.92	-2.51	-1.57	-2.35
0.10	-1.63	-1.47	-1.52	-1.95	-1.55	-1.88	-1.74	-1.27	-1.51	-2.13	-1.19	-1.86
0.25	-0.91	-0.77	-0.79	-1.22	-0.79	-0.99	-1.08	-0.63	-0.82	-1.51	-0.56	-1.02
0.50	-0.18	-0.04	-0.03	-0.44	-0.03	-0.06	-0.39	0.06	-0.10	-0.81	0.09	-0.13
0.75	0.48	0.61	0.64	0.25	0.47	0.64	0.30	0.73	0.62	-0.12	0.69	0.71
0.90	1.03	1.13	1.19	0.71	0.82	1.07	0.90	1.33	1.27	0.48	1.19	1.43
0.95	1.33	1.42	1.50	0.92	0.99	1.27	1.24	1.65	1.61	0.80	1.47	1.79
$\sigma_\eta^2=1$												
0.05	-2.20	-1.98	-2.13	-2.68	-2.16	-2.83	-2.19	-1.62	-2.03	-2.74	-1.47	-3.18
0.10	-1.74	-1.52	-1.64	-2.20	-1.64	-2.30	-1.83	-1.25	-1.62	-2.40	-1.11	-2.66
0.25	-1.04	-0.81	-0.88	-1.52	-0.89	-1.46	-1.18	-0.61	-0.91	-1.79	-0.51	-1.52
0.50	-0.27	-0.05	-0.05	-0.74	-0.12	-0.39	-0.49	0.07	-0.13	-1.10	0.12	-0.28
0.75	0.40	0.57	0.66	-0.01	0.27	0.55	0.20	0.73	0.62	-0.40	0.65	0.85
0.90	0.91	1.00	1.16	0.46	0.56	0.95	0.79	1.29	1.27	0.20	1.05	1.69
0.95	1.17	1.23	1.41	0.65	0.71	1.17	1.11	1.61	1.63	0.49	1.27	2.10

Table 3B
Model 1: Robust estimates
Quantiles of the t statistics

	$T=4$						$T=7$					
	$\alpha=0.5$			$\alpha=0.8$			$\alpha=0.5$			$\alpha=0.8$		
	GMM2	SNM2	LIM2	GMM2	SNM2	LIM2	GMM2	SNM2	LIM2	GMM2	SNM2	LIM2
$\sigma_\eta^2=0$												
0.05	-2.04	-1.97	-1.91	-2.25	-2.12	-2.07	-2.49	-2.24	-2.34	-2.74	-2.24	-2.45
0.10	-1.61	-1.54	-1.47	-1.80	-1.62	-1.55	-2.01	-1.73	-1.82	-2.28	-1.79	-1.90
0.25	-0.87	-0.78	-0.73	-1.00	-0.80	-0.75	-1.22	-0.91	-0.92	-1.47	-0.94	-0.92
0.50	-0.11	0.01	0.06	-0.22	0.02	0.05	-0.33	0.00	0.08	-0.57	-0.03	0.09
0.75	0.58	0.71	0.76	0.45	0.72	0.73	0.56	0.91	1.05	0.28	0.85	1.06
0.90	1.18	1.32	1.35	1.00	1.28	1.26	1.30	1.67	1.89	1.03	1.62	1.89
0.95	1.54	1.69	1.71	1.30	1.61	1.55	1.76	2.12	2.42	1.46	2.05	2.37
$\sigma_\eta^2=0.2$												
0.05	-2.15	-2.08	-2.00	-2.68	-2.71	-2.48	-2.62	-2.31	-2.42	-3.28	-2.53	-2.98
0.10	-1.71	-1.62	-1.55	-2.15	-2.02	-1.84	-2.11	-1.79	-1.86	-2.73	-1.97	-2.22
0.25	-0.93	-0.83	-0.76	-1.28	-1.01	-0.88	-1.30	-0.93	-0.95	-1.88	-1.05	-1.11
0.50	-0.17	0.02	0.05	-0.43	-0.05	0.04	-0.41	-0.02	0.06	-0.97	-0.11	0.05
0.75	0.54	0.71	0.77	0.29	0.75	0.73	0.45	0.87	1.04	-0.05	0.81	1.15
0.90	1.13	1.31	1.34	0.77	1.32	1.15	1.24	1.68	1.90	0.70	1.60	2.01
0.95	1.44	1.65	1.66	0.98	1.76	1.37	1.69	2.13	2.44	1.13	2.06	2.46
$\sigma_\eta^2=1$												
0.05	-2.36	-2.35	-2.26	-3.17	-4.44	-3.01	-2.76	-2.41	-2.55	-3.82	-3.10	-3.72
0.10	-1.83	-1.78	-1.67	-2.58	-3.22	-2.26	-2.27	-1.88	-1.96	-3.26	-2.37	-2.77
0.25	-1.09	-0.95	-0.82	-1.68	-1.67	-1.14	-1.44	-0.98	-1.01	-2.35	-1.31	-1.39
0.50	-0.25	-0.05	0.03	-0.78	-0.33	0.00	-0.56	-0.05	0.03	-1.37	-0.19	0.00
0.75	0.46	0.73	0.77	0.01	0.70	0.70	0.32	0.87	1.07	-0.43	0.82	1.26
0.90	0.98	1.31	1.30	0.50	1.51	1.11	1.09	1.66	1.94	0.35	1.68	2.14
0.95	1.28	1.63	1.56	0.70	2.52	1.40	1.51	2.11	2.46	0.76	2.14	2.59

Notes to Tables 3A and 3B:

10,000 replications. $N=100$, $\sigma_v^2=1$.

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 4A
Employment equations
Robust estimates from the UK sample

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

Independent variables	Model A GMM2	Model A SNM2	Model A LIM2	Model A Indirect GMM2[1]	Model B GMM2	Model B SNM2	Model B LIM2
$\Delta n_{i(t-1)}$	0.800	1.596	1.900	1.214	0.825	2.186	0.836
	(0.048)	(0.105)	(0.173)		(0.056)	(0.216)	(0.060)
$\Delta n_{i(t-2)}$	-0.116	-0.384	0.105	-0.282	-0.074	-0.455	0.344
	(0.021)	(0.045)	(0.053)		(0.020)	(0.077)	(0.038)
Δw_{it}	-0.640	-1.897	0.507	-4.638			
	(0.054)	(0.160)	(0.224)				
$\Delta w_{i(t-1)}$	0.564	2.138	0.487	1.567	0.431	2.841	0.615
	(0.066)	(0.142)	(0.222)		(0.076)	(0.312)	(0.080)
Δk_{it}	0.219	0.238	-1.353	0.604			
	(0.051)	(0.089)	(0.198)				
$\Delta k_{i(t-1)}$					-0.077	-0.787	-0.235
					(0.045)	(0.126)	(0.049)
Δys_{it}	0.890	1.747	0.674	3.105			
	(0.098)	(0.204)	(0.228)				
$\Delta ys_{i(t-1)}$	-0.874	-2.897	-0.006	-4.101	-0.115	-2.438	-0.427
	(0.105)	(0.229)	(0.312)		(0.100)	(0.358)	(0.112)
$\Delta ys_{i(t-2)}$					0.095	1.511	0.126
					(0.091)	(0.266)	(0.101)
Sargan test (df)	63.0 (50)	67.1 (50)	44.5 (50)	62.8 (50)	68.3 (51)	66.5 (51)	57.8 (51)
R^2 's for IVs:							
$\Delta n_{i(t-1)}$	0.271				0.269		
$\Delta w_{it)}$	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 4B
Employment equations
Robust estimates from the UK sample

Dependent variable: Δn_{it}	Sample period: 1979 – 1984 (140 companies)					
	AR(2) Models					
Independent variables	GMM2	SNM2	LIM2	GMM2	SNM2	LIM2
$\Delta n_{i(t-1)}$	0.691	1.635	1.412	0.320	0.827	0.092
	(0.051)	(0.074)	(0.067)	(0.053)	(0.065)	(0.047)
$\Delta n_{i(t-2)}$	-0.114	-0.439	-0.348	0.022	-0.094	0.218
	(0.026)	(0.039)	(0.025)	(0.022)	(0.032)	(0.019)
$\Delta w_{i(t-1)}$	0.598	1.958	0.297			
	(0.070)	(0.095)	(0.073)			
$\Delta w_{i(t-2)}$	0.013	-0.075	-0.163			
	(0.036)	(0.053)	(0.041)			
Sargan test (df)	65.9 (50)	71.3 (50)	48.8 (50)	32.8 (25)	31.3 (25)	31.7 (25)
R^2's for IVs:						
$\Delta n_{i(t-1)}$	0.216			0.152		

Notes to Tables 4A and 4B:
(i) Time dummies are included in all equations.
(ii) Asymptotic standard errors robust to heteroskedasticity are reported in parentheses.
(iii) 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.
(iv) 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 three columns also lags of wages dated $(t-2)$ and earlier.
(v) 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.

Table 5
VAR estimates for employment and wage equations
from the Spanish sample

Sample period: 1983 – 1990 (738 companies)

Independent variables	GMM2	SNM2	LIM2	GMM2	SNM2	LIM2
	"Model 1" restrictions					
Δn_{it} Equation						
$\Delta n_{i(t-1)}$	0.842	1.087	1.004	0.748	0.813	0.832
	(0.669;1.015)	(0.894;1.280)	(0.830;1.178)	(0.575;0.921)	(0.636;0.988)	(0.661;1.002)
	[0.712;1.209]	[0.959;1.485]		[0.505;0.976]	[0.629;1.092]	
$\Delta n_{i(t-2)}$	-0.003	-0.074	-0.049	0.038	0.030	0.027
	(-0.060;0.054)	(-0.140;-0.008)	(-0.110;0.012)	(-0.005;0.081)	(-0.015;0.075)	(-0.018;0.072)
	[-0.146;0.028]	[-0.244;-0.039]		[-0.027;0.084]	[-0.046;0.073]	
$\Delta w_{i(t-1)}$	0.078	0.222	0.177			
	(-0.086;0.242)	(0.046;0.398)	(0.016;0.338)			
	[-0.006;0.412]	[0.124;0.624]				
$\Delta w_{i(t-2)}$	-0.053	-0.074	-0.068			
	(-0.102;-0.004)	(-0.127;-0.021)	(-0.121;-0.015)			
	[-0.116;-0.002]	[-0.138;-0.020]				
Sargan test (df)	36.9 (36)	37.2 (36)	35.5 (36)	14.4 (18)	13.5 (18)	13.0 (18)
R^2's for IVs:						
$\Delta n_{i(t-1)}$	0.033			0.022		
$\Delta w_{i(t-1)}$	0.031					
Δw_{it} Equation						
$\Delta w_{i(t-1)}$	0.178	0.228	0.063	0.178	0.228	0.063
	(-0.042;0.398)	(-0.008;0.464)	(-0.176;0.302)	(-0.042;0.398)	(-0.008;0.464)	(-0.176;0.302)
	[-0.075;0.405]	[-0.100;0.482]		[-0.144;0.429]	[-0.232;0.519]	
$\Delta w_{i(t-2)}$	-0.012	-0.002	-0.039	-0.012	-0.002	-0.039
	(-0.081;0.049)	(-0.066;0.062)	(-0.102;0.024)	(-0.081;0.049)	(-0.066;0.062)	(-0.102;0.024)
	[-0.076;0.042]	[-0.077;0.052]		[-0.089;0.045]	[-0.100;0.060]	
Sargan test (df)	12.7 (18)	12.9 (18)	12.2 (18)	12.7 (18)	12.9 (18)	12.2 (18)
R^2's for IVs:						
$\Delta w_{i(t-1)}$	0.019					

Notes to Table 5:

(i) Time dummies are included in all equations.

(ii) The instrument set for all the employment equations includes lags of employment dated $(t-2)$ and earlier, and for those in the first three columns also lags of wages dated $(t-2)$ and earlier. The instrument set for the wage equation includes lags of wages dated $(t-2)$ and earlier.

(iii) 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.

(iv) 95% asymptotic confidence intervals based on heteroskedasticity-robust standard errors in parentheses; 95% moment-restricted bootstrap confidence intervals in brackets. The bootstrap confidence intervals for the equations in the first three columns are based on a distribution that satisfies a larger set of moment conditions than those in the last three columns. The reason is that the former include lagged wages as instruments for the employment equation, which are absent from the latter.

Table 6
VAR estimates for employment and wage equations from the Spanish sample
Sample period: 1983 – 1990 (738 companies)

Independent variables	"Model 2" restrictions GMM2	SNM2	LIM2
Δn_{it} *Equation*			
$\Delta n_{i(t-1)}$	1.163	1.208	1.624
	(1.112;1.214)	(1.137;1.279)	(1.424;1.824)
	[1.132;1.218]	[1.143;1.229]	
$\Delta n_{i(t-2)}$	-0.135	-0.142	-0.160
	(-0.172;-0.098)	(-0.185;-0.099)	(-0.231;-0.089)
	[-0.197;-0.108]	[-0.206;-0.117]	
$\Delta w_{i(t-1)}$	0.121	0.116	0.058
	(0.086;0.156)	(0.077;0.155)	(-0.001;0.117)
	[0.091;0.161]	[0.094;0.164]	
$\Delta w_{i(t-2)}$	-0.132	-0.151	-0.242
	(-0.171;-0.093)	(-0.194;-0.108)	(-0.313;-0.171)
	[-0.173;-0.101]	[-0.177;-0.101]	
Sargan test (df)	80.1 (48)	69.1 (48)	50.3 (48)
Δw_{it} *Equation*			
$\Delta w_{i(t-1)}$	0.854	0.873	0.869
	(0.815;0.893)	(0.834;0.912)	(0.828;0.911)
	[0.825;0.902]	[0.828;0.905]	
$\Delta w_{i(t-2)}$	0.152	0.138	0.141
	(0.105;0.199)	(0.089;0.187)	(0.090;0.192)
	[0.099;0.186]	[0.094;0.183]	
Sargan test (df)	71.4 (24)	72.2 (24)	71.4 (24)

Notes to Table 6:
(i) Time dummies are included in all equations.

(ii) The instrument set for the employment equations includes lags of employment and wages dated $(t-2)$ and earlier for errors in first differences, and lags of employment and wages in first differences dated $(t-1)$ for errors in levels. The instrument set for the wage equations is similar, but excludes lagged employment in levels and first differences.

(iii) GMM2 and SNM2 are two-step estimates based on one-step GMM residuals that use all the orthogonality restrictions from Model 2, and the inverse of the second moments of the instruments as the weighting matrix.

(iv) 95% asymptotic confidence intervals based on heteroskedasticity-robust standard errors in parentheses; 95% moment-restricted bootstrap confidence intervals in brackets.

Table 7
VAR estimates for employment and wage equations
from the Spanish sample
Random sample containing 200 companies

Sample period: 1983 – 1990

Independent variables	GMM2	SNM2	LIM2	GMM2	SNM2	LIM2
Δn_{it} *Equation*						
$\Delta n_{i(t-1)}$	0.788	1.160	1.002	0.441	0.815	1.517
	(0.610;0.966)	(0.888;1.432)	(0.777;1.227)	(0.167;0.715)	(0.509;1.121)	(1.081;1.952)
	[0.528;1.248]	[0.932;1.903]		[0.217;0.983]	[0.424;1.214]	
$\Delta n_{i(t-2)}$	-0.042	-0.206	-0.181	0.063	0.003	-0.170
	(-0.109;0.025)	(-0.306;-0.106)	(-0.271;-0.091)	(0.002;0.124)	(-0.062;0.069)	(-0.268;-0.072)
	[-0.265;-0.008]	[-0.567;-0.120]		[-0.060;0.120]	[-0.138;0.090]	
$\Delta w_{i(t-1)}$	0.337	0.650	0.675			
	(0.151;0.523)	(0.371;0.929)	(0.452;0.898)			
	[0.099;0.680]	[0.300;1.048]				
$\Delta w_{i(t-2)}$	0.001	-0.040	-0.018			
	(-0.065;0.067)	(-0.120;0.040)	(-0.098;0.062)			
	[-0.150;0.059]	[-0.261;0.006]				
Sargan test (df)	30.2 (36)	23.0 (36)	24.8 (36)	23.3 (18)	24.3 (18)	16.5 (18)
R^2 's for IVs:						
$\Delta n_{i(t-1)}$	0.064			0.040		
$\Delta w_{i(t-1)}$	0.080					
Δw_{it} *Equation*						
$\Delta w_{i(t-1)}$	-0.612	-1.198	-1.246	-0.612	-1.198	-1.246
	(-0.984;-0.240)	(-1.442;-0.953)	(-1.509;-0.983)	(-0.984;-0.240)	(-1.442;-0.953)	(-1.509;-0.983)
	[-0.962;0.359]	[-3.512;2.492]		[-0.954;0.402]	[-4.893;4.932]	
$\Delta w_{i(t-2)}$	-0.120	-0.270	-0.231	-0.120	-0.270	-0.231
	(-0.231;-0.009)	(-0.349;-0.191)	(-0.319;-0.143)	(-0.231;-0.009)	(-0.349;-0.191)	(-0.319;-0.143)
	[-0.232;0.102]	[-0.627;0.348]		[-0.239;0.183]	[-1.202;0.993]	
Sargan test (df)	17.3 (18)	11.0 (18)	9.3 (18)	17.3 (18)	11.0 (18)	9.3 (18)
R^2 's for IVs:						
$\Delta w_{i(t-1)}$	0.023					

See Notes to Table 5.

Table 8
Monte Carlo simulations for the AR(2) model for employment
$\alpha_1 = 0.813$, $\alpha_2 = 0.030$, $\gamma = 0.777$,
$\sigma_\eta^2 = 0.038$, $\sigma_v^2 = 0.01$

		$N = 738$		$N = 200$	
		GMM2	SNM2	GMM2	SNM2
Summary of estimates					
α_1					
	median	0.72	0.82	0.55	0.82
	% bias	12.0	0.3	32.2	0.8
	iqr	0.14	0.15	0.27	0.28
	iq80	0.28	0.29	0.56	0.61
	mae	0.11	0.08	0.26	0.14
α_2					
	median	0.01	0.03	-0.02	0.02
	% bias	64.6	7.0	163.3	35.4
	iqr	0.04	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 statistics					
α_1					
	0.10	-2.44	-1.37	-3.61	-1.62
	0.25	-1.75	-0.74	-2.77	-0.82
	0.50	-1.01	0.02	-1.84	0.04
	0.75	-0.25	0.77	-0.97	0.81
	0.90	0.41	1.33	-0.21	1.42
α_2					
	0.10	-2.22	-1.55	-2.93	-1.92
	0.25	-1.48	-0.82	-2.16	-1.05
	0.50	-0.78	-0.08	-1.26	-0.24
	0.75	-0.01	0.60	-0.45	0.60
	0.90	0.62	1.19	0.17	1.08

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.