ELSEVIER

Journal of Econometrics 68 (1995) 29–51

JOURNAL OF Econometrics

Another look at the instrumental variable estimation of error-components models

Manuel Arellano*[,a], Olympia Bover[b]

[a]CEMFI, 28014 Madrid, Spain

[b]Research Department, Bank of Spain, 28014 Madrid, Spain

Abstract

This article develops a framework for efficient IV estimators of random effects models with information in levels which can accommodate predetermined variables. Our formulation clarifies the relationship between the existing estimators and the role of transformations in panel data models. We characterize the valid transformations for relevant models and show that optimal estimators are invariant to the transformation used to remove individual effects. We present an alternative transformation for models with predetermined instruments which preserves the orthogonality among the errors. Finally, we consider models with predetermined variables that have constant correlation with the effects and illustrate their importance with simulations.

Key words: Dynamic panel data; Predetermined instrumental variables; Orthogonal deviations; Unrestricted covariance matrix; Unit roots

JEL classification: C23

1. Introduction

The static error components model with both time-invariant and time-varying explanatory variables allowing for the correlation of some of these

* Corresponding author.

An earlier version of this paper was presented at the World Congress of the Econometric Society in Barcelona, August 1990, and at the Canadian Econometrics Study Group in Guelph, September 1990. It circulated as Centre for Economic Performance Discussion Paper No. 7 (London School of Economics, August 1990). We thank Steve Bond, David Card, Gary Chamberlain, Franco Peracchi, and specially Peter Schmidt for useful comments. All remaining errors are our own.

SSDI 030440769401642 D

variables with the unobservable individual effects was first considered by Hausman and Taylor (1981) – hereafter HT. Bhargava and Sargan (1983) – hereafter BS – studied the estimation of dynamic error components models, and also considered a model which contained a lagged dependent variable and allowed for correlation between some of the regressors and the effects. Subsequently, Amemiya and MaCurdy (1986) – hereafter AM – and Breusch, Mizon, and Schmidt (1989) – hereafter BMS – developed alternative instrumental variable (IV) estimators of the HT model that are more efficient than the original HT estimator. On the other hand, Anderson and Hsiao (1982), Holtz-Eakin, Newey, and Rosen (1988), and Arellano and Bond (1991), amongst others, considered the estimation of models with predetermined but no strictly exogenous variables by IV methods using lagged values of the predetermined variables as instruments for the equations in first differences. In these models it is usually maintained that all the explanatory variables are potentially correlated with the individual effects and therefore only estimators based on deviations of the original observations can be consistent. However, if there are instruments available that are not correlated with the effects, the levels of the variables contain information concerning the parameters of interest which if exploited could improve, sometimes crucially, the efficiency of the resulting estimates. In addition, this information in the levels may be sufficient to identify the coefficients of time-invariant explanatory variables that are correlated with the effects.

The purpose of this paper is to develop a framework for efficient IV estimators with information in levels which is capable of accommodating models with lagged dependent variables and other predetermined variables. In Section 2 we present a generalised method of moments formulation of HT, AM, and BMS-like estimators. Each particular model gives rise to a set of orthogonality restrictions on which estimation is to be based. We follow Amemiya and MaCurdy in exploiting transformations of the original equations in order to obtain convenient expressions of these restrictions. However, we formulate the matrices of instruments as block-diagonal matrices with as many blocks as the total number of time periods. In this way we can show that the optimal estimators are invariant to the choice of transformation. Another advantage of proceeding in this way is that we can obtain HT, AM, and BMS estimators with nonstandard or unrestricted covariance matrix without having to specify the appropriate GLS transformation and subsequent changes to the instrument set to avoid inconsistencies. As noted by AM, since different instruments are only valid for subsets of equations, GLS transformations are sensitive in this context: a particular IV matrix that is valid for some GLS transformation of the model may be invalid under a different GLS transformation. By specifying an IV matrix that effectively lists all the individual moment restrictions available we avoid this problem. We also calculate the Fisher information bound for the parameters of a conditional moment specification of the model in order to assess the efficiency of the class of GMM estimators formulated in Section 2.

Section 3 shows how the previous framework can easily accommodate dynamic models, and other models with predetermined variables and information in levels. We discuss an IV estimator which is asymptotically equivalent to the limited information maximum likelihood (LIML) estimator with unrestricted covariance matrix and correlated exogenous variables of BS. This clarifies the relationship between HT/AM/BMS and BS. We also extend these estimators to include lags of predetermined variables as additional instruments. We characterise the class of valid transformations in this context and show the invariance of the optimal estimators to a particular choice of transformation. We argue that a computationally convenient transformation for these models is forward orthogonal deviations. A closely related transformation has been used by Hayashi and Sims (1983) for time series models. This transformation leads to simple expressions of the estimators in terms of the vectors of instruments corresponding to individual time periods, and so it avoids the need to operate with the full block-diagonal IV matrix which may have an excessively large number of columns. Section 3 also formulates a GMM estimator for a general model with predetermined variables and information in levels.

Section 4 considers a model with predetermined variables that have constant correlation with the individual effects. As an illustration of the potential of these constraints, we report Monte Carlo simulations of IV estimators of a first-order stationary autoregression with random effects that exploit the orthogonality restrictions in levels. An estimator that only uses the restrictions in first differences is also simulated for comparisons. The section concludes with some remarks on the usefulness of predetermined variables that have constant correlation with the effects for the testing of unit roots in short panels. Finally, Section 5 contains the conclusions of the paper.

2. A method of moments formulation of Hausman–Taylor and related estimators with unrestricted covariance matrix

Let us consider the model

$$\begin{aligned} y_{it} &= \beta' x_{it} + \gamma' f_i + u_{it}, \qquad t = 1, \dots, T, \quad i = 1, \dots, N, \\ u_{it} &= \eta_i + v_{it}, \\ \mathrm{E}(v_{it} | x_{i1}, \dots, x_{iT}, f_i, \eta_i) &= 0. \end{aligned} \tag{1}$$

So that the variables x_{it} and f_i are assumed to be strictly exogenous given the unobservable individual effect η_i. Under standard conditions, this assumption identifies β but not γ. The identification of γ is based on the following assumption:

$$\mathrm{E}(\eta_i | x_{1i1}, \dots, x_{1iT}, f_{1i}) = 0, \tag{2}$$

where we are using the partitions $x_{it} = (x'_{1it}, x'_{2it})'$ and $f_i = (f'_{1i}, f'_{2i})'$. Throughout, T is small and N is large. This model can be regarded as an intermediate case between the 'fixed effects' model in which all the explanatory variables are potentially correlated with the effects and therefore only estimators based on deviations of the observations can be consistent, and the standard uncorrelated 'random effects' model in which $x_{1it} = x_{it}$ and $f_{1i} = f_i$.

It is convenient to re-write (1) in the form

$$y_i = W_i \delta + u_i, \tag{3}$$

where $y_i = (y_{i1}, \dots, y_{iT})'$, $u_i = (u_{i1}, \dots, u_{iT})'$, $\delta = (\beta', \gamma')'$, $W_i = (X_i | \iota f_i')$, $X_i = (x_{i1}, \dots, x_{iT})'$ and ι is a $T \times 1$ vector of ones. Below, we also make use of the notation $\bar{x}_i' = T^{-1}\iota' X_i = (\bar{x}'_{1i}, \bar{x}'_{2i})$ and the vectors $v_i = (v_{i1}, \dots, v_{iT})'$, $x_i = (x'_{i1}, \dots, x'_{iT})'$, and $w_i = (x_i', f_i')'$.

In general, the matrix $\mathrm{E}(u_i u_i' | w_i)$ will be unrestricted and depend on w_i:

$$\mathrm{E}(u_i u_i' | w_i) = \mathrm{E}(v_i v_i' | w_i) + \mathrm{E}(\eta_i^2 | w_i) \iota\iota' = \Omega(w_i).$$

However, here we emphasize two cases with cross-sectional homoskedasticity in the sense that $\mathrm{E}(u_i u_i' | w_i) = \mathrm{E}(u_i u_i')$.[1] Firstly, the case of a constant unrestricted Ω, which allows for the possibility of autocorrelation and time series heteroskedasticity of arbitrary form in the v_{it}. Secondly, the traditional error components specification given by $\Omega = \sigma^2 I_T + \sigma_\eta^2 \iota\iota'$, where I_T is the identity matrix of order T.

We then transform the system of T equations using a nonsingular $T \times T$ transformation matrix,

$$H = \begin{bmatrix} K \\ T^{-1}\iota' \end{bmatrix},$$

where K is any $(T-1) \times T$ matrix of rank $(T-1)$ such that $K\iota = 0$. For example, K could be the first difference operator or the first $(T-1)$ rows of the within-group operator. The transformed errors are given by

$$u_i^+ = H u_i = \begin{bmatrix} K u_i \\ \bar{u}_i \end{bmatrix}. \tag{4}$$

This class of transformations performs a decomposition between 'within-group' and 'between-group' variation which is helpful in order to implement

[1] We assume that $\mathrm{E}(\eta_i) = \mathrm{E}(\mathrm{E}(\eta_i | w_i)) = 0$. Notice that, provided the model contains a constant term, there is no loss of generality in this assumption. Thus, although it is always true that $\mathrm{E}(u_i u_i') = \mathrm{var}(u_i) = \mathrm{var}(v_i) + \mathrm{var}(\eta_i)\iota\iota'$, in general $\mathrm{E}(u_i u_i' | w_i)$ and $\mathrm{var}(u_i | w_i) = \mathrm{var}\,(y_i | w_i)$ will differ as follows:

$$\mathrm{E}(u_i u_i' | w_i) = \mathrm{var}(u_i | w_i) + (\mathrm{E}(\eta_i | w_i))^2 \iota\iota'.$$

orthogonality restrictions implied by the model. Specifically, since the first $(T-1)$ errors do not contain η_i, all exogenous variables (as well as nonlinear functions of those variables) are valid instruments for the first $(T-1)$ equations. Then, if m_i denotes a vector of a subset of variables of w_i (or linear combinations of those variables) assumed to be uncorrelated in levels and such that $\dim(m_i) \geqslant \dim(\gamma)$, a valid IV matrix for the complete transformed system is

$$Z_i = \begin{bmatrix} w_i' & & & 0 \\ & \ddots & & \\ & & w_i' & \\ 0 & & & m_i' \end{bmatrix}.$$

We can now write down the optimal GMM estimator of δ with constant Ω based on the moment equations,

$$\mathrm{E}(Z_i' H u_i) = 0,$$

which is given by

$$\hat{\delta} = [W'\bar{H}'Z(Z'\bar{H}\bar{\Omega}\bar{H}'Z)^{-1}Z'\bar{H}W]^{-1}W'\bar{H}'Z(Z'\bar{H}\bar{\Omega}\bar{H}'Z)^{-1}Z'\bar{H}y, \tag{5}$$

where $W = (W_1' \dots W_N')'$, $y = (y_1' \dots y_N')'$, $Z = (Z_1' \dots Z_N')'$, $\bar{H} = I_N \otimes H$, and $\bar{\Omega} = I_N \otimes \Omega$. In practice, the covariance matrix of the transformed system $\Omega^+ = H\Omega H'$ will be replaced by a consistent estimator. An unrestricted estimator of Ω^+ takes the form

$$\hat{\Omega}^+ = \frac{1}{N}\sum_{i=1}^{N} \hat{u}_i^+ \hat{u}_i^{+\prime},$$

where the $\hat{u}_i^+$ are residuals based on consistent preliminary estimates. Alternatively, we consider a restricted estimate $\tilde{\Omega}^+ = H\tilde{\Omega}H'$ with $\tilde{\Omega} = \tilde{\sigma}^2 I_T + \tilde{\sigma}_\eta^2 \imath\imath'$, where $\tilde{\sigma}^2$ and $\tilde{\sigma}_\eta^2$ denote consistent estimates of σ^2 and σ_η^2.

The estimator of HT is $\hat{\delta}$ with $\tilde{\Omega}^+$ and

$$m_i = (f_{1i}', \bar{x}_{1i}')',$$

whereas the estimator of AM is $\hat{\delta}$ with $\tilde{\Omega}^+$ and

$$m_i = (f_{1i}'\ x_{1i1}' \dots x_{1iT}')'.$$

BS and BMS also exploited the additional moment restrictions that arise if it is assumed that the correlation between x_{2it} and η_i is constant over time. In this case, the deviations from time means $\tilde{x}_{2it} = x_{2it} - \bar{x}_{2i}$ are valid instruments for the last equation of the transformed system. A stronger conditional expectation version of this assumption along the lines of (2) is

$$\mathrm{E}(\eta_i | x_{1i}, f_{1i}, \tilde{x}_{2i}) = 0. \tag{6}$$

Setting

$$m_i = (f'_{1i}\ x'_{1i1}\ \dots\ x'_{1iT}\ \tilde{x}'_{2i2}\ \dots\ \tilde{x}'_{2iT})'$$

and using $\tilde{\Omega}^+$, $\hat{\delta}$ gives the estimator of BMS. Moreover, if all variables are uncorrelated with the effects, we can set $m_i = w_i$ in which case $\hat{\delta}$ with $\tilde{\Omega}^+$ becomes the GLS estimator of Balestra and Nerlove (1966). On the other hand, if all variables are correlated with the effects, the levels equation drops out, the coefficients γ are unidentified and estimation of β is based on $\mathrm{E}(Z'_{di}Ku_i) = 0$ with $Z_{di} = I_{(T-1)} \otimes w'_i$. In the case of restricted Ω since $K\Omega K' = \sigma^2 KK'$, letting $\bar{K} = I_N \otimes K$, $X = (X'_1\ \dots\ X'_N)'$, and $Z_d = (Z'_{d1}\ \dots\ Z'_{dN})'$, the resulting estimator is

$$\hat{\beta} = [X'\bar{K}'Z_d(Z'_d\bar{K}\bar{K}'Z_d)^{-1}Z'_d\bar{K}X]^{-1}X'\bar{K}'Z_d(Z'_d\bar{K}\bar{K}'Z_d)^{-1}Z'_d\bar{K}y,$$

which can be shown to coincide with the within-group estimator.

It is interesting to notice that having chosen a block-diagonal form for Z_i, $\hat{\delta}$ is invariant to the choice of transformation K. To prove this assertion we can use the following simple fact in GMM estimation. The optimal estimator of θ based on $\mathrm{E}[\xi_i(\theta)] = 0$ minimizes

$$s = (\Sigma_i\xi_i)'\hat{A}^{-1}(\Sigma_i\xi_i),$$

where $\hat{A}$ is a consistent estimator of $\mathrm{E}(\xi_i\xi'_i)$. If we now consider $\xi_i^* = F\xi_i$ where F is a nonsingular transformation matrix, it turns out that the optimal estimator of θ based on ξ_i^* that minimizes

$$s^* = (\Sigma_i\xi_i^*)'\hat{A}^{*-1}(\Sigma_i\xi_i^*)$$

is numerically the same estimator as the one based on ξ_i provided that $\hat{A}^* = F\hat{A}F'$ since $s = s^*$. In our case

$$\xi_i = Z'_iHu_i = \begin{bmatrix} Ku_i \otimes w_i \\ \bar{u}_i m_i \end{bmatrix} = \begin{bmatrix} K \otimes I_{p_1} & 0 \\ 0 & T^{-1}\iota' \otimes I_{p_2} \end{bmatrix} \begin{bmatrix} \mathrm{vec}(u_i w'_i) \\ \mathrm{vec}(u_i m'_i) \end{bmatrix}, \qquad (7)$$

where p_1 and p_2 are the number of elements in w_i and m_i, respectively, and the vec operator stacks the elements of a matrix by rows. Suppose that an alternative transformation $H^* = (K^{*\prime}\ k^*\iota')'$ is used. Letting $K^* = \Phi K$ and $k^* = \varphi T^{-1}$ we can write

$$\xi_i^* = Z'_iH^*u_i = F\xi_i,$$

where

$$F = \begin{bmatrix} \Phi \otimes I_{p_1} & 0 \\ 0 & \varphi I_{p_2} \end{bmatrix}.$$

So that any valid transformation leads to the same estimator.

This is useful because in this way we can obtain HT, AM, and BMS-like estimators easily with various specifications for Ω without having to specify the appropriate $\Omega^{-1/2}$ transformation and subsequent changes to the instrument set to avoid inconsistencies. It also provides a natural framework to extend the HT-type of estimators to cases where there are predetermined variables as we shall see below.

If Ω^+ is estimated as $\tilde{\Omega}^+$, straightforward manipulations reveal that (5) can be written as

$$\hat{\delta} = [\Sigma_i W_i' Q W_i + \tilde{\theta}^2 T \Sigma_i \bar{w}_i m_i' (\Sigma_i m_i m_i')^{-1} \Sigma_i m_i \bar{w}_i']^{-1}$$
$$\times [\Sigma_i W_i' Q y_i + \tilde{\theta}^2 T \Sigma_i \bar{w}_i m_i' (\Sigma_i m_i m_i')^{-1} \Sigma_i m_i \bar{y}_i]^{-1}, \qquad (8)$$

which produces more familiar expressions of the HT, AM, and BMS estimators for the corresponding choices of m_i (details available in the Appendix).[2] In this expression Q is the within-group operator:

$$Q = I_T - \iota\iota'/T = K'(KK')^{-1}K,$$
$$\bar{w}_i = W_i'\iota/T \quad \text{and} \quad \tilde{\theta}^2 = \tilde{\sigma}^2/(\tilde{\sigma}^2 + T\tilde{\sigma}_\eta^2).$$

As explained in the Appendix, in this case it is possible to simplify the form of Z_i without changing the estimator.

The obvious advantage of the formulation (5) is that if we replace the error components estimator $\tilde{\Omega}^+$ by an unrestricted estimator $\hat{\Omega}^+$, we obtain alternative HT, AM, or BMS-type estimators which are as efficient asymptotically as the versions in (8) when $\mathrm{E}(v_i v_i') = \sigma^2 I_T$ and strictly more efficient when $\mathrm{E}(v_i v_i') \neq \sigma^2 I_T$. Moreover, with cross-sectional heteroskedasticity further efficiency can be achieved using a GMM estimator of the type discussed by Chamberlain (1982), Hansen (1982), and White (1982) which would replace the term $(\Sigma_i Z_i' \Omega^+ Z_i)$ in (5) by a term of the form $(\Sigma_i Z_i' \hat{u}_i^+ \hat{u}_i^{+\prime} Z_i)$.

The efficiency bound for δ

In order to assess the efficiency of the class of GMM estimators given in (5), it is useful to compare the inverse of the asymptotic variance matrix of $\hat{\delta}$ with the Fisher information bound for δ based on the conditional moment restrictions (1) and (2). Chamberlain (1992a), using a specification that includes (1) as a special case, shows that the bound for β based on (1) is identical to the bound for β based on the conditional moment restriction

$$\mathrm{E}\,(K(y_i - X_i\beta) \mid w_{1i}, w_{2i}) = 0, \qquad (9)$$

[2] A derivation of the estimators of HT, AM, and BMS as GMM estimators, in the case that noise is iid, has been obtained independently by Ahn and Schmidt (1995).

where $w_{ji} = (x'_{ji}, f'_{ji})'$, $j = 1, 2$, so that $(w'_{1i}, w'_{2i})'$ is just a permutation of w_i. In addition, (2) can be written as

$$E(y_i - W_i\delta \mid w_{1i}) = 0. \tag{10}$$

The Fisher information bound for δ based on (9) and (10) can be obtained as an application of Theorem 1 of Chamberlain (1992b) for sequential conditional moment restrictions.[3] The bound will be the sum of the bounds corresponding to each of the conditional moments (see the Appendix for the details):

$$E(Ku_i \mid w_i) = 0, \tag{11}$$

$$E[(\iota'\Omega_i^{-1}\iota)^{-1}\,\iota'\Omega_i^{-1}u_i \mid w_{1i}] = 0, \tag{12}$$

where $\Omega_i = \Omega(w_i)$. Direct application of Chamberlain's theorem gives the following expression for the bound:

$$J = E\big(W_i'K'(K\Omega_iK')^{-1}KW_i + [E((\iota'\Omega_i^{-1}\iota)^{-1} \mid w_{1i})]^{-1} \times E(\bar{\bar{w}}_i \mid w_{1i})\,E(\bar{\bar{w}}_i' \mid w_{1i})\big), \tag{13}$$

where $\bar{\bar{w}}_i' = q_i'W_i$ and $q_i' = (\iota'\Omega_i^{-1}\iota)^{-1}\iota'\Omega_i^{-1}$. With a constant unrestricted Ω, J becomes

$$J = E\big(W_i'K'(K\Omega K')^{-1}KW_i + (\iota'\Omega^{-1}\iota)E(W_i' \mid w_{1i})qq'E(W_i \mid w_{1i})\big). \tag{14}$$

None of the GMM estimators of this section will attain the bound even in the absence of cross-sectional heteroskedasticity, since $E(W_i \mid w_{1i})$ could be a nonlinear function of w_{1i}. However, if $E(W_i \mid w_{1i})$ is linear, we have

$$E(W_i'q \mid w_{1i}) = E(\bar{\bar{w}}_i \mid w_{1i}) = E(\bar{\bar{w}}_i \mid w_{1i}')\,[E(w_{1i}w_{1i}')]^{-1}w_{1i}$$

and

$$J = E\big(W_i'K'(K\Omega K')^{-1}KW_i + (\iota'\Omega^{-1}\iota)E(\bar{\bar{w}}_i w_{1i}')\,[E(w_{1i}w_{1i}')]^{-1}\,E(w_{1i}\bar{\bar{w}}_i')\big). \tag{15}$$

Finally, if $\Omega = \sigma^2 I_T + \sigma_\eta^2\iota\iota'$, we have that $q = T^{-1}\iota$, $\iota'\Omega^{-1}\iota = \theta^2T/\sigma^2$, and $K'(K\Omega K')^{-1}K = (1/\sigma^2)Q$, so that

$$J = (1/\sigma^2)\,E\big(W_i'QW_i + \theta^2T\,E(\bar{w}_i w_{1i}')\,[E(w_{1i}w_{1i}')]^{-1}E(w_{1i}\bar{w}_i')\big), \tag{16}$$

which equals the inverse of the asymptotic covariance matrix of the AM estimator.

Notice that the assumptions of linearity of $E(W_i \mid w_{1i})$ and of a constant error components structure for Ω would imply further conditional moment restrictions that may lower the information bound for δ. Here, we merely particularize

[3] This theorem applies to the case where the joint distribution of the data is multinomial but it could be extended to a general distribution by using the approximation argument of Chamberlain (1987).

the bound for δ based on (11) and (12) to the case where these additional restrictions happen to occur in the population but are not used in the calculation of the bound.

We now turn to show that the inverse of the asymptotic covariance matrix of the GMM estimator given in (5) with unrestricted Ω and the AM choice of instruments for the average equation $m_i = w_{1i}$, V^{-1} say, coincides with the information bound given in (15). Under standard regularity conditions

$$V^{-1} = \mathrm{E}(W_i'H'Z_i)\,[\mathrm{E}(Z_i'H\Omega H'Z_i)]^{-1}\,\mathrm{E}(Z_i'HW_i).$$

On the other hand, since KX_i equals the block $I_{T-1} \otimes w_i'$ of Z_i multiplied by a constant selection matrix, after straightforward manipulations (15) can be written as

$$J = \mathrm{E}(W_i'H^{*\prime}Z_i)\,[\mathrm{E}(Z_i'H^*\Omega H^{*\prime}Z_i)]^{-1}\,\mathrm{E}(Z_i'H^*W_i),$$

with

$$H^* = \begin{bmatrix} K \\ \imath'\Omega^{-1} \end{bmatrix}.$$

We prove that $V^{-1} = J$ by showing that

$$Z_i'H^* = F(Z_i'H), \tag{17}$$

where F is a nonsingular matrix of constants. Firstly, notice that $H^* = \Phi H$ with

$$\Phi = \begin{bmatrix} I_{T-1} & 0 \\ \imath'\Omega^{-1}K'(KK')^{-1} & \imath'\Omega^{-1}\imath \end{bmatrix}.$$

Next, we consider a permutation of the columns of $Z_i = Z_i^* P'$ such that

$$Z_i^* = \begin{bmatrix} I_T \otimes w_{1i}' & \begin{matrix} I_{T-1} \otimes w_{2i}' \\ 0 \end{matrix} \end{bmatrix}.$$

Hence notice

$$Z_i^{*\prime} H = \begin{bmatrix} (I_T \otimes w_{1i})H \\ (I_T \otimes w_{2i})K \end{bmatrix} = \begin{bmatrix} H \otimes I & 0 \\ 0 & K \otimes I \end{bmatrix} \begin{bmatrix} I_T \otimes w_{1i} \\ I_T \otimes w_{2i} \end{bmatrix},$$

and similarly

$$Z_i^{*\prime} H^* = \begin{bmatrix} H^* \otimes I & 0 \\ 0 & K \otimes I \end{bmatrix} \begin{bmatrix} I_T \otimes w_{1i} \\ I_T \otimes w_{2i} \end{bmatrix},$$

which proves (17) with

$$F = P\begin{bmatrix} \Phi \otimes I & 0 \\ 0 & I \end{bmatrix} P^{-1}.$$

Remark that the previous result depends crucially on the fact that the variables in the instrument set for the between-group equation are linear combinations of the instruments used for the within-group equations. This means, for example, that in the block $I_{T-1} \otimes w_i'$ of Z_i we cannot replace w_i with vec(KX_i), hence excluding f_{1i} and $\bar{x}_{1i}$, without altering the estimator and its asymptotic variance, contrary to the situation in the case that the noise is iid as explained in the Appendix. Nevertheless, some of the instruments used for the within-group equations are redundant in the sense that their omission would not alter the GMM estimator. Specifically, the situation is that for HT, AM, and BMS choices of m_i with unrestricted Ω, the submatrix $I_{T-1} \otimes w_i'$ of Z_i could be replaced by $I_{T-1} \otimes ([\text{vec}(KX_i)]', f_{1i}', \bar{x}_{1i}')$ leaving the estimator unaltered. Another remark is that the difference between the asymptotic covariance matrix of the HT estimator with unrestricted Ω and V will be a nonnegative matrix, except if $\mathrm{E}^*(W_i | w_{1i})$ coincides with $\mathrm{E}(W_i | \bar{x}_{1i}, f_{1i})$ where E^* denotes a best linear predictor, in which case the two estimators have the same asymptotic variances.

One last remark concerns BMS-type estimators. Clearly, our analysis can be repeated for BS/BMS models by using the conditional moment restriction (6) in place of (2). The previous discussion applies provided the vectors of conditioning variables are suitably redefined. In this case, the vector of IV for the between-group equation is $m_i = (f_{1i}' \; x_{1i}' \; \tilde{x}_{2i2}' \; \dots \; \tilde{x}_{2iT}')'$. However, for this choice of m_i the rows of KX_i are linear combinations of m_i. This means that the same instrument set is valid for all the equations and we can use $Z_i = I_T \otimes m_i'$ without altering the estimator. The consequence is that the transformation is unnecessary and the estimator can be obtained by simple application of three-stage least squares (3SLS) to the original system of equations using m_i as the vector of instruments for all equations:

$$\hat{\delta} = [\Sigma_i(W_i \otimes m_i)'(\hat{\Omega} \otimes \Sigma_i m_i m_i')^{-1}\Sigma_i(W_i \otimes m_i)]^{-1}\Sigma_i(W_i \otimes m_i)' \times (\hat{\Omega} \otimes \Sigma_i m_i m_i')^{-1}\Sigma_i(y_i \otimes m_i). \tag{18}$$

3. Models with predetermined variables and a useful transformation

We begin by considering a model of the type given in (1) with the addition of the dependent variable lagged one period:

$$y_{it} = \alpha y_{i(t-1)} + \beta' x_{it} + \gamma' f_i + u_{it}, \qquad u_{it} = \eta_i + v_{it}, \tag{19}$$

$$\mathrm{E}(v_{it} | x_{i0}, \dots, x_{iT}, f_i, \eta_i) = 0. \tag{20}$$

Assuming that $t = 0$ is observed and redefining the symbols in (3) as $\delta = (\alpha, \beta', \gamma')'$ and $W_i = (y_{i(-1)}, X_i, \imath f_i')$ with $y_{i(-1)} = (y_{i0}, \ldots, y_{i(T-1)})'$, the expression given in (5) remains a consistent GMM estimator of δ for this model, provided there are enough valid instruments to ensure identification. The form of the IV matrix Z_i is the same as in Section 2, adjusting for the fact that $t = 0$ is now observed, so that $w_i = (x_{i0}', \ldots, x_{iT}', f_i')$ (notice the exclusion of $y_{i(-1)}$ despite its presence in W_i). The same range of choices for m_i are available depending on the assumptions concerning the dependence between η_i and subsets of the explanatory variables.

In particular, if $m_i = (f_{1i}', x_{1i}', \tilde{x}_{2i1}', \ldots, \tilde{x}_{2iT}')$, in view of the reasons explained for BMS-type cases above, the resulting estimator coincides with 3SLS and is therefore asymptotically equivalent to the LIML procedure with Ω unrestricted developed by Bhargava and Sargan. BS obtained their estimator as an application of subsystem LIML to the T equations (19), having completed the system with the reduced form equations,

$$y_{i0} = \pi_0' m_i + \varepsilon_{i0}, \quad f_{2i} = \Pi_1 m_i + \varepsilon_{i1}, \quad \bar{x}_{2i} = \Pi_2 m_i + \varepsilon_{i2},$$

and the identities

$$x_{2it} = \tilde{x}_{2it} + \bar{x}_{2i}, \qquad t = 0, \ldots, T-1.$$

It is well-known that subsystem LIML is asymptotically equivalent to subsystem 3SLS when Ω is unrestricted.

As in the static model, one polar case is the uncorrelated random effects specification with $\mathrm{E}(\eta_i | x_i, f_i) = 0$, so that $m_i = w_i$, which corresponds to the basic model of BS. At the other end, η_i would be potentially correlated with all explanatory variables and there would be no instruments for the levels equation, which would drop out. This corresponds to the model and the 3SLS estimator discussed by Chamberlain (1984, pp. 1266–1267).

In the previous model, regardless of the existence of individual effects, unrestricted serial correlation in v_{it} implies that $y_{i(t-1)}$ is an endogenous variable. A different model, in which $y_{i(t-1)}$ is a predetermined variable given η_i, replaces (20) by the following assumption:

$$\mathrm{E}(v_{it} | x_i, f_i, \eta_i, y_{i0}, \ldots, y_{i(t-1)}) = 0. \tag{21}$$

Notice that (21) implies lack of serial dependence in the sense that $\mathrm{E}(v_{it} | v_{i1} \ldots v_{i(t-1)}) = 0$. Orthogonality restrictions implied by this model can be easily incorporated in an estimator of the form of (5) provided that the transformation matrix K is upper-triangular in addition to the previous requirements. In effect, with lack of autocorrelation in v_{it} and K upper-triangular it turns out that the transformed error in the equation for period t is independent of η_i and $(v_{i1}, \ldots, v_{i(t-1)})$ so that $(y_{i0}, \ldots, y_{i(t-1)})$ are additional valid instruments for this equation. Hence giving rise to the following Z_i matrix:

$$Z_i = \begin{bmatrix} w_i' \; y_{i0} & & & & 0 \\ & w_i' \; y_{i0} \; y_{i1} & & & \\ & & \ddots & & \\ & & & w_i' \; y_{i0} \cdots y_{i(T-2)} & \\ 0 & & & & m_i' \end{bmatrix}. \tag{22}$$

Estimators that rely on these types of restrictions have been discussed by Anderson and Hsiao (1982), Holtz-Eakin, Newey, and Rosen (1988), and Arellano and Bond (1991). These authors transformed the data using first differences and disregarded the levels in the absence of valid instruments for this equation (Arellano and Bond (p. 280) did, however, present a discussion of models with predetermined and strictly exogenous variables that contain information in levels). Further discussion of these models is contained in Ahn and Schmidt (1995), who exploit the additional quadratic moment restrictions implied by lack of serial correlation and the restrictions derived from the assumption of homoskedasticity. A model may contain predetermined variables other than lags of the dependent variable, but their treatment would be similar to the one described for $y_{(t-1)}$. Moreover, it is often the case that instruments arising from assumptions on predetermined variables and lack of autocorrelation are the only ones available in the model, so that sequential moment restrictions like (21) become crucial for the identification of the parameters of interest.

As in the previous section, the GMM estimator (5) that uses (22) as the matrix of instruments is invariant to the choice of K provided K satisfies the required conditions. This is an example of a more general result: let z_{is} be the $r_s \times 1$ vector of instruments that are valid in the transformed equations for periods $s, s+1, \ldots, (T-1)$ [for example, in (22) $z_{i1} = (w_i' \; y_{i0})'$ and $z_{is} = y_{i(s-1)}$ for $s > 1$], and let K_s be the $(T-s-1) \times T$ submatrix that results when the first s rows of K are eliminated. Then the moment restrictions available for estimation are

$$\mathrm{E}(\xi_i) = \mathrm{E}\begin{bmatrix} Ku_i \otimes z_{i1} \\ K_1 u_i \otimes z_{i2} \\ \vdots \\ K_{T-2} u_i \otimes z_{i(T-1)} \\ T^{-1} \iota' u_i m_i \end{bmatrix} = 0. \tag{23}$$

Since ξ_i can be written as

$$\xi_i = \begin{bmatrix} K \otimes I_{r_1} & & & 0 \\ & \ddots & & \\ & & K_{T-2} \otimes I_{r_{T-1}} & \\ 0 & & & T^{-1}\iota' \otimes I_{p_2} \end{bmatrix} \begin{bmatrix} \mathrm{vec}(u_i z'_{1i}) \\ \vdots \\ \mathrm{vec}(u_i z'_{i(T-1)}) \\ \mathrm{vec}(u_i m'_i) \end{bmatrix},$$

it turns out that for any other valid $K^* = \Phi K$ the resulting ξ_i^* will be of the form $F\xi_i$ with F having the following block-diagonal structure:

$$F = \mathrm{diag}(\Phi \otimes I_{r_1}, \ldots, \Phi_1 \otimes I_{r_2}, \ldots, \Phi_{T-2} \otimes I_{r_{T-1}}, I_{p_2}),$$

where $K_s^* = \Phi_s K_s$. As a consequence, all the estimators of the form (5) with K upper-triangular, $K\iota = 0$ and Z_i given by

$$Z_i = \begin{bmatrix} z'_{i1} & & & & 0 \\ & z'_{i1}\, z'_{i2} & & & \\ & & \ddots & & \\ & & & z'_{i1} \cdots z'_{i(T-1)} & \\ 0 & & & & m'_i \end{bmatrix},$$

are identical.

However, as pointed out by Schmidt, Ahn, and Wyhowski (1992) who stress the point that filtering does not improve efficiency of estimation if all available instruments are used, this does not mean that filtering is useless, since in practice it may not be desirable to use all of the available instruments for computational reasons or if their number is excessive for the actual sample size, given the finite-sample properties of the estimators.

Orthogonal deviations

An alternative to first differencing which is very useful in the context of models with predetermined variables is the following Helmert's transformation:

$$u_{it}^* = c_t\left[u_{it} - \frac{1}{(T-t)}(u_{i(t+1)} + \cdots + u_{iT})\right], \qquad t = 1, \ldots, T-1, \tag{24}$$

where $c_t^2 = (T-t)/(T-t+1)$. That is, to each of the first (T − 1) observations we subtract the mean of the remaining future observations available in the sample. The weighting c_t is introduced to equalise the variances. The choice of

K that produces this transformation is the forward orthogonal deviations operator:

$$A = \operatorname{diag}\left[\frac{T-1}{T}, \ldots, \frac{1}{2}\right]^{1/2} \times$$

$$\begin{bmatrix} 1 & -(T-1)^{-1} & -(T-1)^{-1} & \cdots & -(T-1)^{-1} & -(T-1)^{-1} & -(T-1)^{-1} \\ 0 & 1 & -(T-2)^{-1} & \cdots & -(T-2)^{-1} & -(T-2)^{-1} & -(T-2)^{-1} \\ \vdots & \vdots & \vdots & & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 1 & -\frac{1}{2} & -\frac{1}{2} \\ 0 & 0 & 0 & \cdots & 0 & 1 & -1 \end{bmatrix} \tag{25}$$

which clearly has rows whose elements add up to zero (so that the permanent effects are eliminated) and is upper-triangular (so that lags of predetermined variables are valid instruments in the transformed equations). In addition, it preserves the orthogonality among the transformed errors – if the original v_{it} are not autocorrelated and have constant variance, so are the transformed errors, and indeed it can be verified by direct multiplication that $AA' = I_{(T-1)}$ and $A'A = Q$. Hence, $A = (KK')^{-1/2}K$ for any upper-triangular K, so that for example transforming by A can be regarded as doing first differences to eliminate the effects plus a GLS transformation to remove the serial correlation induced by differencing.

A useful feature of this transformation when $\Omega = \sigma^2 I + \sigma_\eta^2 \iota\iota'$ is that since it diagonalises $H\Omega H'$ it is possible to calculate $\hat{\delta}$ in the following way:

$$\begin{aligned} \hat{\delta} = \Bigg[& \sum_{t=1}^{T-1} (\Sigma_i w_{it}^* z_{it}')(\Sigma_i z_{it} z_{it}')^{-1}(\Sigma_i z_{it} w_{it}^{*\prime}) \\ & + \tilde{\theta}^2 T(\Sigma_i \bar{w}_i m_i')(\Sigma_i m_i m_i')^{-1}(\Sigma_i m_i \bar{w}_i') \Bigg]^{-1} \\ \times \Bigg[& \sum_{t=1}^{T-1} (\Sigma_i w_{it}^* z_{it}')(\Sigma_i z_{it} z_{it}')^{-1}(\Sigma_i z_{it} y_{it}^*) \\ & + \tilde{\theta}^2 T(\Sigma_i \bar{w}_i m_i')(\Sigma_i m_i m_i')^{-1}(\Sigma_1 m_i \bar{y}_i) \Bigg], \end{aligned} \tag{26}$$

where w_{it}^* is the tth row of $W_i^* = AW_i$ and y_{it}^* is the tth element of Ay_i. This is of importance in practice because if Z_i has a large number of columns it may be difficult to compute expression (5) directly.

Finally, notice that since $A'A = Q$, the OLS regression of y_{it}^* on x_{it}^* (that is, least squares applied to the first $(T-1)$ equations of the system) will give the within-group estimator, whereas OLS applied to the complete system of T equations with $H = (A', \tilde{\theta}T^{-1/2}\imath)'$ will give the GLS estimator.

A general model with predetermined variables and information in levels

Combining together the various ingredients that have appeared so far, the form of a general model with predetermined variables and information in levels is as follows:

$$y_{it} = w_{it}'\delta + \eta_i + v_{it}, \tag{27}$$

$$\mathrm{E}(v_{it} \mid x_{i1}, \ldots, x_{iT}, f_i, p_{i1}, \ldots, p_{it}, \eta_i) = 0, \tag{28}$$

$$\mathrm{E}(\eta_i \mid x_{1i1}, \ldots, x_{1iT}, f_{1i}, p_{1i1}, \ldots, p_{1iT}) = 0. \tag{29}$$

The vector of right-hand-side variables w_{it} may include lags of y_{it}, time-invariant variables f_i, plus other strictly exogenous, predetermined, or endogenous variables. The variables f_i, x_{it}, and p_{it} refer to time-invariant, strictly exogenous, and predetermined variables, respectively. For each category we introduce the partitions $f_i = (f_{1i}', f_{2i}')'$, $x_{it} = (x_{1it}', x_{2it}')'$, and $p_{it} = (p_{1it}', p_{2it}')'$, with the first subsets denoting the variables that are uncorrelated to η_i according to (29).

Notice that (28) and (29) imply that

$$\mathrm{E}(y_{it} - w_{it}'\delta \mid x_{1i1}, \ldots, x_{1iT}, f_{1i}, p_{1i1}, \ldots, p_{1it}) = 0, \tag{30}$$

so that in the presence of p_{1it} variables there are different instruments available for different equations in levels, what precludes the use of the average equation in constructing GMM estimators.[4] Following Arellano and Bond (1991), we can define a $(2T-1) \times T$ transformation $H = (K', I_T)'$ and

$$Z_i = \begin{bmatrix} Z_{di} & 0 \\ 0 & Z_{li} \end{bmatrix}, \tag{31}$$

where Z_{di} is block-diagonal and has the tth block given by $(x_{1i}', \ldots, x_{iT}', f_i', p_{i1}', \ldots, p_{i(t-1)}')$, which are the instruments available for the tth equation transformed by K. The matrix Z_{li} is also block-diagonal and will contain the instruments available for the equations in levels. In principle, in the equation

[4] Chamberlain (1992b) obtained the Fisher information bound for δ in a model similar to (27) and (28), with the exclusion of (29). However, the addition of (29) breaks the sequential moment structure of the problem with the implication that Chamberlain's results are not directly applicable to this case.

for period t, the vector of valid instruments is $(x'_{1i1}, \dots, x'_{1iT}, f'_{1i}, p'_{1i1}, \dots, p'_{1it})$. However, given Z_{di}, some of these moment restrictions will be redundant. To see this, taking K to be the first difference operator without loss of generality, notice that

$$\mathrm{E}(u_{it}p_{i(t-s)}) = \sum_{j=0}^{s-1} \mathrm{E}(\Delta u_{i(t-j)}p_{i(t-s)}) + \mathrm{E}(u_{i(t-s)}p_{i(t-s)}). \tag{32}$$

Therefore, we specify Z_{li} as

$$Z_{li} = \begin{bmatrix} x'_{1i1} \cdots x'_{1iT} \;\; f'_{1i} \;\; p'_{1i1} & & 0 \\ & p'_{1i2} & \\ & & \ddots & \\ 0 & & & p'_{1iT} \end{bmatrix}. \tag{33}$$

We can construct optimal GMM estimators of δ based on the moment equations

$$\mathrm{E}(Z'_i H u_i) = \mathrm{E}\begin{bmatrix} Z'_{di}Ku_i \\ Z'_{li}u_i \end{bmatrix} = 0. \tag{34}$$

So we are replacing the 'between-group' errors in (4) by the complete set of errors in levels in addition to those transformed by K. Individual equations in levels rather than an average equation are now required since we have different instruments valid for different equations in levels. The next section presents a model of special interest which contains predetermined variables that are valid instruments in the equations in levels.

Note that the estimators of HT, AM, and BMS can also be written in this way, for example selecting $Z_{di} = I_{T-1} \otimes w'_i$ and $Z_{li} = I_T \otimes m'_i$ and using expression (5). The only modification that (5) requires is the replacement of the inverse of $\Sigma_i Z'_i \Omega^+ Z_i$ by a Moore–Penrose generalised inverse, since this matrix will be singular due to repetitions of the same moments.

4. Additional moment restrictions using predetermined variables

The models of BS and BMS included strictly exogenous variables that had constant correlation with the individual effects. That is, variables such that

$$\mathrm{E}(x_{it}\eta_i) = \mathrm{E}(x_{is}\eta_i), \tag{35}$$

$$\mathrm{E}(x_{it}v_{is}) = 0, \tag{36}$$

for all t and s. Here we consider a model with predetermined variables that have constant correlation with the effects. These variables will therefore satisfy (35) for all t and s, but (36) will only be true for $t \leqslant s$.[5] This type of restrictions could be justified on the grounds of stationarity, and in many instances its validity or otherwise can be regarded as an empirical issue. Moreover, in models without strictly exogenous variables, like vector autoregressions and some rational expectations models, these additional restrictions may play a crucial role in substantially improving the precision of the estimates, especially when T is very small.

Estimation can proceed as a special case of the general model with predetermined variables and information in levels discussed in the previous section. Suppose for simplicity of presentation that in (1) all the x_{it} are predetermined variables that satisfy (35) and that all the f_i are correlated with η_i. Therefore in the equation in first differences for period t $(x_{i1}, x_{i2}, \dots, x_{i(t-1)})$ are valid instruments while in the equation in levels $(\Delta x_{i2}, \dots, \Delta x_{it})$ are valid. Some of these moment restrictions are redundant. To see this note that

$$\mathrm{E}(u_{it}\Delta x_{i(t-1)}) - \mathrm{E}(u_{i(t-1)}\Delta x_{i(t-1)}) = \mathrm{E}(\Delta u_{it} x_{i(t-1)}) - \mathrm{E}(\Delta u_{it} x_{i(t-2)}),$$

so that given three restrictions that equate three of these four terms to zero the equality of the fourth term to zero is redundant. Thus, given the instruments for the first difference equations, (35) contributes the additional constraints

$$\mathrm{E}(u_{it}\Delta x_{it}) = 0, \qquad t = 1, \dots, T.$$

Redefining H as the $2(T-1)\times T$ transformation $H = (K', I_0')'$ with $I_0 = (0 | I_{T-1})$ and choosing $Z_{di} = \mathrm{diag}[x_{i1}', \dots, (x_{i1}' \dots x_{i(T-1)}')]$ and $Z_{li} = \mathrm{diag}(\Delta x_{i2}', \dots, \Delta x_{iT}')$, we can construct optimal GMM estimators of β and γ based on the moment equations $\mathrm{E}(Z'Hu_i) = 0$.

Monte Carlo results

Finally, we have carried out simulations concerning a well-known simple model: a first-order autoregression with random effects observed three time periods. The purpose of the experiments is to illustrate the potential of exploiting moment restrictions in levels equations using predetermined variables in first differences. For each experiment we generated 1000 samples of N independent observations of (y_{i0}, y_{i1}, y_{i2}) from the process

$$y_{i0} = (1-\alpha)^{-1}\eta_i + (1-\alpha^2)^{-1/2} v_{i0},$$

$$y_{i1} = \alpha y_{i0} + \eta_i + v_{i1},$$

$$y_{i2} = \alpha y_{i1} + \eta_i + v_{i2}, \qquad i = 1, \dots, N,$$

with $v_i = (v_{i0}\ v_{i1}\ v_{i2})' \sim \mathrm{N}(0, \mathrm{I})$ and $\eta_i \sim \mathrm{N}(0, \sigma_\eta^2)$ independent of v_i.

---

[5] A stronger conditional mean version of this situation would assume that x_{it} can be written as $x_{it} = g(\eta_i) + x_{it}^+$ and $\mathrm{E}(\eta_i | x_{i1}^+ \dots x_{iT}^+) = 0$. If x_{it} is strictly exogenous, $\mathrm{E}(v_{it} | x_{i1} \dots x_{iT}) = 0$, whereas if it is predetermined, $\mathrm{E}(v_{it} | x_{i1} \dots x_{it}) = 0$.

The Anderson–Hsiao estimator of α is based on the restriction

$$\mathrm{E}[(\Delta y_{i2} - \alpha \Delta y_{i1}) y_{i0}] = 0, \tag{37}$$

and is given by

$$\hat{\alpha}_{AH} = \frac{\Sigma_i y_{i0} \Delta y_{i2}}{\Sigma_i y_{i0} \Delta y_{i1}}. \tag{38}$$

Moreover we are also interested in exploiting the levels restriction

$$\mathrm{E}[(y_{i2} - \alpha y_{i1}) \Delta y_{i1}] = 0, \tag{39}$$

so that we consider the system of two equations

$$\begin{bmatrix} \Delta y_{i2} \\ y_{i2} \end{bmatrix} = \alpha \begin{bmatrix} \Delta y_{i1} \\ y_{i1} \end{bmatrix} + \begin{bmatrix} \Delta u_{i2} \\ u_{i2} \end{bmatrix},$$

with the matrix of instruments

$$Z_i = \begin{bmatrix} y_{i0} & 0 \\ 0 & \Delta y_{i1} \end{bmatrix}.$$

Let $y_{it}^{+} = (\Delta y_{it}\ y_{it})'$. We simulate two estimators of the form

$$\hat{\alpha}_L = \frac{(\Sigma_i y_{i1}^{+\prime} Z_i) A_N (\Sigma_i Z_i' y_{i2}^{+})}{(\Sigma_i y_{i1}^{+\prime} Z_i) A_N (\Sigma_i Z_i' y_{i1}^{+})}. \tag{40}$$

The one-step estimator $\hat{\alpha}_{L1}$ sets $A_N = (\Sigma_i Z_i' Z_i)^{-1}$, while the two-step estimator $\hat{\alpha}_{L2}$ uses $A_N = (\Sigma_i Z_i' \hat{u}_i^{+} \hat{u}_i^{+\prime} Z_i)^{-1}$ where $\hat{u}_i^{+} = y_{i2}^{+} - \hat{\alpha}_{L1} y_{i1}^{+}$.

Table 1 reports Monte Carlo means and standard deviations of the three estimators for $\alpha = 0.2, 0.5, 0.8$, $\sigma_\eta^2 = 0, 0.2, 1$, and $N = 100, 500$. σ^2 is kept equal to unity for all the experiments (with $\sigma_\eta^2 = 0$ all three estimators are invariant to the value of σ^2). With $\sigma_\eta^2 = 0.2$, the variation due to the permanent effect represents 23, 37.5, and 57 percent of the total variance of y_{i0} for $\alpha = 0.2, 0.5, 0.8$, respectively. While for $\sigma_\eta^2 = 1$ the corresponding percentages of variation are 60, 75, and 90.

As can be seen in the table, *L1* and *L2* always outperform *AH* both in terms of having a smaller standard deviation and a smaller bias. The gap in precision between *AH* and the estimators that also use the restrictions in levels widens for larger values of σ_η^2 and α. With $N = 100$ and $\alpha = 0.8$, *AH* is a useless estimator whereas *L1* and *L2* behave reasonably well. Even with $N = 500$, for $\alpha = 0.8$, the standard deviation of *AH* is twice that of *L1* and *L2* with $\sigma_\eta^2 = 0$, three times bigger with $\sigma_\eta^2 = 0.2$, and one hundred times bigger with $\sigma_\eta^2 = 1$! The same pattern still applies to $\alpha = 0.5$, with large efficiency gains and reductions in biases obtained by using the restriction in levels. On the other hand, there is not

Table 1
Means and standard deviations of the estimators, 1000 replications

	$\alpha = 0.2$			$\alpha = 0.5$			$\alpha = 0.8$		
	AH	*L1*	*L2*	*AH*	*L1*	*L2*	*AH*	*L1*	*L2*
$N = 100$									
$\sigma_\eta^2 = 0$									
Mean	0.2315	0.2147	0.2018	0.5571	0.5095	0.4953	0.9683	0.7841	0.7795
S.d.	0.1852	0.1460	0.1776	0.2827	0.1729	0.1720	0.8203	0.2027	0.2109
$\sigma_\eta^2 = 0.2$									
Mean	0.2353	0.2128	0.2009	0.5814	0.5001	0.4884	1.3701	0.7496	0.7482
S.d.	0.2134	0.1572	0.1556	0.4088	0.1895	0.1900	17.1953	0.2516	0.2667
$\sigma_\eta^2 = 1$									
Mean	0.2588	0.2065	0.2011	0.7980	0.4762	0.4748	0.0390	0.7526	0.7574
S.d.	0.3389	0.1948	0.1898	2.9516	0.2431	0.2409	15.9819	0.3309	0.3727
$N = 500$									
$\sigma_\eta^2 = 0$									
Mean	0.2056	0.2038	0.2011	0.5097	0.5031	0.4999	0.8248	0.7976	0.7984
S.d.	0.0768	0.0635	0.0622	0.1093	0.0765	0.0734	0.1949	0.0900	0.0870
$\sigma_\eta^2 = 0.2$									
Mean	0.2059	0.2041	0.2012	0.5120	0.5022	0.4983	0.8596	0.7887	0.7886
S.d.	0.0856	0.0695	0.0689	0.1356	0.0864	0.0849	0.3660	0.1105	0.1153
$\sigma_\eta^2 = 1$									
Mean	0.2089	0.2040	0.2019	0.5262	0.4963	0.4917	1.8560	0.7597	0.7600
S.d.	0.1189	0.0906	0.0891	0.2262	0.1155	0.1145	21.1516	0.1775	0.1813

Each sample consists of N independent observations of (y_{i0}, y_{i1}, y_{i2}) generated from the process:

$$y_{i0} = (1-\alpha)^{-1}\eta_i + (1-\alpha^2)^{-1/2}v_{i0}, \qquad y_{i1} = \alpha y_{i0} + \eta_i + v_{i1}, \qquad y_{i2} = \alpha y_{i1} + \eta_i + v_{i2},$$

with $v_i = (v_{i0}, v_{i1}, v_{i2})' \sim \mathrm{N}(0, I)$ and $\eta_i \sim \mathrm{N}(0, \sigma_\eta^2)$ independent of v_i.

much difference between the behaviour of *L1* and *L2*. There seems to be a tendency of *L2* to have a smaller standard deviation than *L1* but the gain is negligible.

A specially interesting case is when the coefficient α in

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

is allowed to take the value of unity. If $\eta_i = (1-\alpha)\eta_i^*$, where η_i^* represents the individual specific mean of y_{it} assumed to have a constant variance, with $\alpha = 1$ we have

$$y_{it} = y_{i(t-1)} + v_{it}.$$

The alternative specification of the model with $\alpha = 1$ would be a random walk with an individual drift η_i. In the former case, with $\alpha = 1$, $E(y_{i0} \Delta y_{i1}) = 0$ and as a consequence the Anderson–Hsiao restriction (37) fails to identify α. However, the level restriction (39) still applies and could be exploited in order to test the stationary autoregressive model against the random walk model without drift.

5. Conclusions

Models with predetermined variables for panel data are typically estimated in first differences using instruments in levels. In these models, the absence of information about the parameters of interest in the levels of the variables results in the loss of what sometimes is a very substantial part of the total variation in the data. In this article, we are concerned with panel data models that specify valid instruments for the equations in levels, in addition to those available for the equations in first differences or deviations from individual means. Static models of this kind, but using exclusively strictly exogenous explanatory variables, were first considered by Hausman and Taylor (1981) and, with the addition of a lagged dependent variable, by Bhargava and Sargan (1983). The impact of these models in applied work has been limited, partly due to the difficulty in finding exogenous variables that can be convincingly regarded a priori as being uncorrelated with the individual effects, and partly due to the difficulty in finding strictly exogenous variables at all.

This paper considers models with predetermined instrumental variables that are uncorrelated with the effects. The particular type of variables of this kind that we emphasize are first differences of predetermined variables that have a constant correlation with the effects. A similar assumption for strictly exogenous variables was previously exploited by Bhargava and Sargan (1983) and Breusch, Mizon, and Schmidt (1989). Thus, in addition to using instruments in levels for equations in first differences, we propose to use instruments in first differences for equations in levels. The potential gains in precision from using these constraints are illustrated by means of simulations of alternative estimators of an autoregressive model. Moreover, we also explain how the assumption of constant correlation with the effects can be exploited to test for unit roots in short panels against a stationary autoregressive model.

The paper presents a GMM formulation of Hausman–Taylor (HT) and related estimators with unrestricted covariance matrix, together with a derivation of the information bound for these models. We use this framework to extend HT-type estimators to models with predetermined variables. In doing this we unify a large literature in a coherent way. We propose a GMM estimator for a general model that includes time-invariant, strictly exogenous, and predetermined variables, a subset of which are uncorrelated with the effects. We also show that optimal estimators are invariant to the transformation used to remove the effects. Finally, we propose a new transformation, forward orthogonal deviations, which is a computationally convenient alternative for models with predetermined variables since it preserves the orthogonality among the errors.

Appendix

A.1. GMM formulation of HT, AM, and BMS estimators

Let

$$W_i^+ = HW_i = \begin{bmatrix} KW_i \\ \bar{w}_i' \end{bmatrix} \quad \text{and} \quad Z_i = \begin{bmatrix} Z_{di} & 0 \\ 0 & m_i' \end{bmatrix},$$

so that

$$W_i^{+\prime} Z_i = (W_i' K' Z_{di} \,|\, \bar{w}_i m_i').$$

Now using that $K\iota = 0$ we have $KW_i = (KX_i|0)$ and with $\Omega = \sigma^2 I_T + \sigma_\eta^2 \iota\iota'$,

$$\Omega^+ = H\Omega H' = \begin{bmatrix} K \\ T^{-1}\iota' \end{bmatrix} (\sigma^2 I_T + \sigma_\eta^2 \iota\iota')(K' \; T^{-1}\iota) = \sigma^2 \begin{bmatrix} KK' & 0 \\ 0 & (\theta^2 T)^{-1} \end{bmatrix}.$$

Therefore

$$Z_i' \Omega^+ Z_i = \sigma^2 \begin{bmatrix} Z_{di}' KK' Z_{di} & 0 \\ 0 & (\theta^2 T)^{-1} m_i m_i' \end{bmatrix},$$

and

$$\begin{aligned} & W'\bar{H}'Z(Z'\bar{H}\bar{\Omega}\bar{H}'Z)^{-1} Z'\bar{H}W \\ &= \Sigma_i W_i^{+\prime} Z_i (\Sigma_i Z_i' \Omega^+ Z_i)^{-1} \Sigma_i Z_i' W_i^+ \\ &= \frac{1}{\sigma^2} \left[\begin{bmatrix} M_d & 0 \\ 0 & 0 \end{bmatrix} + \theta^2 T \Sigma_i \bar{w}_i m_i' (\Sigma_i m_i m_i')^{-1} \Sigma_i m_i \bar{w}_i' \right], \end{aligned} \tag{A.1}$$

where

$$M_d = \Sigma_i X_i' K' Z_{di} (\Sigma_i Z_{di}' KK' Z_{di})^{-1} \Sigma_i Z_{di}' K' X_i. \tag{A.2}$$

Now with $Z_{di} = I_{T-1} \otimes w_i'$, M_d equals

$$M_d = \Sigma_i X_i' (I_T \otimes w_i') \left[K'(KK')^{-1} K \otimes (\Sigma_i w_i w_i')^{-1} \right] \Sigma_i (I_T \otimes w_i) X_i.$$

Moreover, since $K'(KK')^{-1}K = Q = A'A$, where A is the orthogonal deviations operator defined in Section 3, and the columns of AX_i are linear combinations of the columns of Z_{di}, M_d becomes

$$M_d = \Sigma_i X_i' A' Z_{di} (\Sigma_i Z_{di}' Z_{di})^{-1} \Sigma_i Z_{di}' A X_i = \Sigma_i X_i' A' A X_i,$$

so that (A.1) equals

$$(1/\sigma^2)\,(\Sigma_i W_i' Q W_i + \theta^2 T \Sigma_i \bar{w}_i m_i' \,(\Sigma_i m_i m_i')^{-1}\, \Sigma_i m_i \bar{w}_i').$$

The derivation of the second term of (8) follows along the same lines.

Note that this result only requires that the columns of AX_i are linear combinations of the columns of Z_{di} provided Z_{di} has the Kronecker structure. Thus, the estimators of the form given in (5) that use $\tilde{\Omega}$ remain unaltered if instead of $Z_{di} = I \otimes w_i'$ we use

$$Z_{di} = I \otimes [\mathrm{vec}(KX_i)]'.$$

In addition, if we choose $K = A$, the block-diagonal specification of Z_{di} could be replaced by simply AX_i without changing the estimator, as apparent from expression (A.2).

A.2. The information bound for Hausman–Taylor models

The model specifies the conditional moment restrictions

$$\mathrm{E}(u_i | w_{1i}) = 0 \quad \text{and} \quad \mathrm{E}(Ku_i | w_{1i}, w_{2i}) = 0.$$

Let us introduce the notation

$$\rho_{1i} = y_i - W_i \delta = u_i,$$

$$\rho_{2i} = K(y_i - W_i \delta) = Ku_i,$$

where $\rho_{ji} = \rho_j(y_i, w_i, \delta)$, $j = 1, 2$. Following Chamberlain (1992b) we consider a forward transformation of ρ_{1i} of the form

$$\rho_{1i}^* = \rho_{1i} - \Gamma(w_i)\,\rho_{2i},$$

which given sequential conditioning ensures that $\mathrm{E}(\rho_{1i}^* | w_{1i}) = 0$.

We wish to choose $\Gamma(w_i)$ such that $\mathrm{E}(\rho_{1i}^* \rho_{2i}' | w_i) = 0$. Since

$$\mathrm{E}(\rho_{1i}^* \rho_{2i}' | w_i) = \Omega_i K' - \Gamma(w_i)\, K \Omega_i K',$$

the condition is satisfied if

$$\Gamma(w_i) = \Omega_i K' (K \Omega_i K')^{-1},$$

where $\Omega_i = \Omega(w_i)$. Thus

$$\rho_{1i}^* = u_i - \Omega_i K' (K \Omega_i K')^{-1} K u_i = \iota(\iota' \Omega_i^{-1} \iota)^{-1}\; \iota' \Omega_i^{-1} u_i.$$

Therefore, the bound for δ will be the sum of the bounds corresponding to each of the conditional moments

$$\mathrm{E}(Ku_i \,|\, w_i) = 0,$$

$$\mathrm{E}[(\iota'\Omega_i^{-1}\iota)^{-1}\iota'\Omega_i^{-1}u_i \,|\, w_{1i}] = 0.$$

References

Ahn, S. and P. Schmidt, 1995, Efficient estimation of models for dynamic panel data, this issue.

Amemiya, T. and T.E. MaCurdy, 1986, Instrumental-variable estimation of an error-components model, Econometrica 54, 869–881.

Anderson, T.W. and C. Hsiao, 1982, Formulation and estimation of dynamic models using panel data, Journal of Econometrics 18, 47–82.

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.

Balestra, P. and M. Nerlove, 1966, Pooling cross section and time series data in the estimation of a dynamic model: The demand for natural gas, Econometrica 34, 585–612.

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

Breusch, T.S., G.E. Mizon, and P. Schmidt, 1989, Efficient estimation using panel data, Econometrica 57, 695–700.

Chamberlain, G., 1982, Multivariate regression models for panel data, Journal of Econometrics 18, 5–46.

Chamberlain, G., 1984, Panel data, in: Z. Griliches and M.D. Intriligator, eds., Handbook of econometrics, Vol. 2 (Elsevier, Amsterdam) 1247–1313.

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

Chamberlain, G., 1992a, Efficiency bounds for semiparametric regression, Econometrica 60, 567–596.

Chamberlain, G., 1992b, Comment: Sequential moment restrictions in panel data, Journal of Business and Economic Statistics 10, 20–26.

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

Hausman, J.A. and W.E. Taylor, 1981, Panel data and unobservable individual effects, Econometrica 49, 1377–1398.

Hayashi, F. and C. Sims, 1983, Nearly efficient estimation of time series models with predetermined, but not exogenous, instruments, Econometrica 51, 783–798.

Holtz-Eakin, D., W. Newey, and H. Rosen, 1988, Estimating vector autoregressions with panel data, Econometrica 56, 1371–1395.

Schmidt, P., S.C. Ahn, and D. Wyhowski, 1992, Comment, Journal of Business and Economic Statistics 10, 10–14.

White, H., 1982, Instrumental variables regression with independent observations, Econometrica 50, 483–499.