ELSEVIER

Journal of Econometrics 68 (1995) 53–78

JOURNAL OF
Econometrics

On bias, inconsistency, and efficiency of various estimators in dynamic panel data models

Jan F. Kiviet

Tinbergen Institute and Faculty of Economics and Econometrics, University of Amsterdam, 1018 WB Amsterdam, The Netherlands

Abstract

When a model for panel data includes lagged dependent explanatory variables, then the habitual estimation procedures are asymptotically valid only when the number of observations in the time dimension (T) gets large. Usually, however, such datasets have substantial sample size in the cross-section dimension (N), whereas T is often a single-digit number. Results on the asymptotic bias ($N \to \infty$) in this situation have been published a decade ago, but, hence far, analytic small sample assessments of the actual bias have not been presented. Here we derive a formula for the bias of the Least-Squares Dummy Variable (LSDV) estimator which has a $O(N^{-1}T^{-3/2})$ approximation error. In a simulation study this is found to be remarkably accurate. Due to the small variance of the LSDV estimator, which is usually much smaller than the variance of consistent (Generalized) Method of Moments estimators, a very efficient procedure results when we remove the bias from the LSDV estimator. The simulations contain results for a particular operational corrected LSDV estimation procedure which in many situations proves to be (much) more efficient than various instrumental variable type estimators.

Key words: Bias correction; Dummy variables estimator; Dynamic panel data model; GMM estimators; Monte Carlo simulation; Unobserved heterogeneity
JEL classification: C13; C23

I am indebted to two anonymous referees and to the editor of this volume for various valuable suggestions. This paper was presented at the 1993 Far-Eastern Meeting of the Econometric Society, 25–27 June, Taipei, Taiwan. It benefited also from the comments by seminar participants at Texas A&M (USA), and the Universities of Amsterdam (The Netherlands), Tartu (Estonia), and Aarhus (Denmark).

SSDI 0304407694016 43 E

1. Introduction

Inference in dynamic regression equations is typically based on large-sample asymptotics, i.e., for $T \to \infty$, where T indicates the number of sample observations of the time-series variables involved. In econometrics the modelling of dynamic relationships usually requires inclusion of lagged dependent explanatory variables. If such dynamic models are estimated from a small (or only moderately large) sample of time-series data, then the standard asymptotic approximations can be very poor (see, for instance, Nankervis and Savin, 1987). The standard approach leads to estimators which (although consistent and asymptotically efficient) are often seriously biased in small samples, and yields test procedures that (although asymptotically valid) are afflicted with an actual type I error probability which differs considerably from the nominal size.

In the case of dynamic panel data models, the asymptotic approximations can be for $T \to \infty$ or for $N \to \infty$ or for both, where N indicates the number of units in each cross-section of the longitudinal sample. In practice T is often very small and N moderate or reasonably large. The accuracy and efficiency of various types of estimators in dynamic error-components models and in dynamic fixed-effects models have been the central issue of a vast number of theoretical and Monte Carlo studies, see *inter alia* Balestra and Nerlove (1966), Nerlove (1971), Maddala (1971), and Arellano and Bond (1991). Here we shall review some of this literature. Next we approximate the small sample bias (for finite N *and* finite T) of the within-group or Least-Squares Dummy Variables (LSDV) estimator for the coefficients of a first-order dynamic panel data model with strongly exogenous explanatory variables and unobserved (random) individual effects. As is well-known, this estimator is *in*consistent for $N \to \infty$ and finite T (see Hsiao, 1986, p.72). Here, we find its small sample bias by applying techniques which have been used so far (as it seems) exclusively in the context of pure time-series dynamic multivariate regression models, see Kiviet and Phillips (1993, 1994). The present results extend findings by Trognon (1978), Nickell (1981), and Sevestre and Trognon (1983, 1985), who derived expressions for the inconsistency (for $N \to \infty$) under alternative situations, such as absence or presence of exogenous explanatory variables, and under various special assumptions concerning the starting values of the dynamic data-generating process.

The approximating formula we derive depends on conditioning variables, but also on the unknown true parameter values. In a Monte Carlo study, where we assess the bias and mean squared error of a whole range of estimators, we find that our approximation (when evaluated at the true parameter values) is very accurate for relevant data generating processes and (small) finite values of N and T. From the same simulations we find, and it also appears from earlier Monte Carlo studies (notably Arellano and Bond, 1991), that the LSDV estimator, although severely biased, has a relatively small dispersion in comparison to various consistent (for $N \to \infty$) estimators. Its standard deviation is often much smaller than obtained for the simple Anderson-Hsiao Instrumental Variables (IV) estimators of the model in first differences and for various Generalized Method of Moments (GMM) estimators. This opens up promising perspectives to exploit an estimate of the LSDV bias for correcting that estimator. If the resulting corrected estimation procedure has minor bias and still a relatively small standard deviation, then, on a mean squared error criterion, this corrected LSDV estimator can be more efficient in finite samples than instrumental variable estimators and possibly even better than (asymptotically efficient) GMM estimators. A particular implementation of the bias assessment is examined, and the resulting operational estimation procedure is compared with other methods. Due to differences in the simulation design, we obtain results which are less favourable for GMM estimators than those presented by Arellano and Bond (1991), and which at the same time illustrate the effectiveness of bias correction of the LSDV estimator.

First, in Section 2, we introduce the model and various estimation techniques, whose small sample behaviour will be examined and compared in later sections. In Section 3 we give an overview of particular relevant earlier results on estimator inconsistency and bias. In Section 4 we derive the approximation to the small sample bias of the LSDV estimator, which has an approximation error of $O(N^{-1}T^{-3/2})$. Section 5 presents Monte Carlo results, and Section 6 concludes.

2. The state dependence model

We consider a dynamic model for panel data with both observed and unobserved exogenous heterogeneity. The unobserved individual effects, which are modelled as random, are possibly correlated with the included exogenous variables, and hence, they cannot simply be dealt with by GLS error-component techniques. The individual effects are to be eliminated by transformations of the model and data, and we consider both taking deviations from group averages (i.e., including dummies) and taking first differences. Both these transformations lead to correlation between the transformed lagged dependent variable and the transformed error term. Nerlove and Balestra (1992) argue that in panel data analysis simpler specifications, for instance static models or models where the unobserved individual effects are uncorrelated with the included explanatory variables, will seldomly be realistic. Also, in cases where the N individuals are a real sample (and not a census), they argue in favour of the random effects specification. See also Hsiao (1985) on this matter.

The present model can be denoted as follows:

$$y_{it} = \gamma y_{i,t-1} + x_{it}'\beta + \eta_i + \varepsilon_{it}, \qquad i = 1, \dots, N, \quad t = 1, \dots, T, \tag{1}$$

where y_{it} is the observation on the dependent variable for unit i at time period t, x_{it} is a $K \times 1$ vector of explanatory variables (the observed heterogeneity) with unknown $K \times 1$ coefficient vector β, η_i denotes the individual effect (unobserved heterogeneity), and ε_{it} is the unobserved disturbance term. The model is dynamic due to the presence of the lagged dependent explanatory variable $y_{i,t-1}$, which has unknown coefficient γ. Probably, the x_{it} vector includes lags of exogenous explanatories too, but such 'finite distributed lag' dynamics does not raise any extra special inference problems, and is therefore left implicit. We make the following further assumptions:

$$\eta_i \sim \mathrm{N}(0, \sigma_\eta^2), \qquad \sigma_\eta^2 \geqslant 0, \tag{2}$$

$$\varepsilon_{it} \sim \mathrm{N}(0, \sigma_\varepsilon^2), \qquad \sigma_\varepsilon^2 > 0, \tag{3}$$

with

$$\begin{array}{lll} \text{(i)} & \mathrm{E}\varepsilon_{it}\varepsilon_{js} = 0, & i \neq j \quad \text{or} \quad t \neq s, \\ \text{(ii)} & \mathrm{E}\eta_i\eta_j = 0, & i \neq j, \\ \text{(iii)} & \mathrm{E}\eta_i\varepsilon_{jt} = 0, & \forall i,j,t, \\ \text{(iv)} & \mathrm{E}x_{it}\varepsilon_{js} = 0, & \forall i,j,t,s, \\ \text{(v)} & \mathrm{E}x_{it}\eta_j = \text{unknown}, & \forall i,j,t. \end{array} \tag{4}$$

The normality assumptions made in (2) and (3) are not essential for all results to follow, but, as usual, they are convenient. The assumptions in (4) state that: (i) the homoscedastic error terms are mutually uncorrelated (over units and over time periods); the unobserved heterogeneity, if present ($\sigma_\eta^2 > 0$), is random and (ii) uncorrelated between individuals and (iii) exogenous (uncorrelated with the error terms); the observed heterogeneity is (iv) strongly exogenous but (v) possibly correlated with the unobserved individual effects. The generality of (v) leads to a situation where appropriate estimators have the same form as in the case of fixed effects, i.e., where the N elements $(\eta_1, \dots, \eta_N)$ are individual coefficients; therefore, error components estimation techniques (see the recent review in Baltagi and Raj, 1992) are ruled out.

In situations where Eq. (1) is a behavioural relationship we usually presuppose dynamic stability, i.e.,

$$|\gamma| < 1. \tag{5}$$

Note that at this stage we do not exclude nonstationarity of the variables x_{it} and y_{it}; they may be integrated, and hence unit roots may occur in univariate models for these variables. By (5) we only restrict relationship (1) in such a way that we suppose the (possibly nonstationary) variables y_{it} and x_{it} to have a stable long-run relationship.

Finally, contrary to the rule in the econometric analysis of pure (nonpanel) time-series relationships, we assume that y_{i0} is random, with

$$\begin{aligned} &\text{(vi)} \quad \mathrm{E}y_{i0}\varepsilon_{jt} = 0, \qquad \forall i,j,t, \\ &\text{(vii)} \quad \mathrm{E}y_{i0}\eta_j = \text{unknown}, \quad \forall ij. \end{aligned} \tag{6}$$

We again refer to arguments given in Nerlove and Balestra (1992), which basically rule out the possibility of fixed y_{i0} in the presence of random individual effects, except for cases where the start of the data-generating process precisely coincides with the start of actual data collection for the sample under study. Otherwise, y_{i0} will naturally be correlated with η_i, but predetermined with respect to the errors ε_{it}.

Under the assumptions made in (2) through (5) model (1) is labelled the 'state dependence model' by Anderson and Hsiao (1982, p. 61). They distinguish it from the model where the presence of a lagged dependent variable is merely due to serial correlation of the error term (common factor dynamics), and they present estimators and their consistency properties (for $N \to \infty$ or $T \to \infty$ or both) under the stationarity assumption:

$$\frac{1}{N \cdot T} \sum_{i=1}^{N} \sum_{t=1}^{T} x_{it} x_{it}' \quad \text{converges to a nonsingular constant matrix.} \tag{7}$$

They also define the latent variable

$$v_{it} = y_{it} - \frac{1}{1-\gamma}\eta_i, \qquad i = 1, \dots, N, \quad t = 0, \dots, T. \tag{8}$$

From substitution in (1) of

$$y_{it} = v_{it} + \frac{1}{1-\gamma}\eta_i, \tag{9}$$

$$v_{it} = \gamma v_{i,t-1} + x_{it}'\beta + \varepsilon_{it}, \qquad i = 1, \dots, N, \quad t = 1, \dots, T, \tag{10}$$

is obtained. From (4), (6), and (8) it follows that

$$\mathrm{E}v_{i0}\varepsilon_{jt} = 0, \qquad \forall i,j,t. \tag{11}$$

For further discussion on the treatment of initial observations see, for instance, Bhargava and Sargan (1983) and Blundell and Smith (1991).

We need to introduce a more succinct notation in order to give simple expressions for the various estimators to be compared. In our derivations to follow it will prove to be convenient to stack the T time period observations for the ith individual's characteristics into T element columns, so we write

$$y_i = \begin{bmatrix} y_{i1} \\ \vdots \\ y_{iT} \end{bmatrix}, \quad y_i^{(-1)} = \begin{bmatrix} y_{i0} \\ \vdots \\ y_{iT-1} \end{bmatrix}, \quad X_i = \begin{bmatrix} x'_{i1} \\ \vdots \\ x'_{iT} \end{bmatrix}, \quad \varepsilon_i = \begin{bmatrix} \varepsilon_{i1} \\ \vdots \\ \varepsilon_{iT} \end{bmatrix}, \quad \iota_T = \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix}. \tag{12}$$

Now (1) can be written as

$$y_i = \gamma y_i^{(-1)} + X_i \beta + \eta_i \iota_T + \varepsilon_i, \tag{13}$$

or as

$$y = \gamma y^{(-1)} + X\beta + (I_N \otimes \iota_T)\eta + \varepsilon, \tag{14}$$

after defining

$$y = \begin{bmatrix} y_1 \\ \vdots \\ y_T \end{bmatrix}, \quad y^{(-1)} = \begin{bmatrix} y_1^{(-1)} \\ \vdots \\ y_N^{(-1)} \end{bmatrix}, \quad X = \begin{bmatrix} X_1 \\ \vdots \\ X_N \end{bmatrix}, \quad \eta = \begin{bmatrix} \eta_1 \\ \vdots \\ \eta_N \end{bmatrix}, \quad \varepsilon = \begin{bmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_N \end{bmatrix}. \tag{15}$$

Here y, $y^{(-1)}$, and ε are $NT \times 1$ vectors, X is a $NT \times K$ matrix, and η is a $N \times 1$ vector. Introducing also

$$W = [y^{(-1)} \vdots X] \quad \text{and} \quad \delta' = (\gamma, \beta'), \tag{16}$$

the model can compactly be written as

$$y = W\delta + (I_N \otimes \iota_T)\eta + \varepsilon. \tag{17}$$

Upon defining

$$A_T = I_T - \frac{1}{T}\iota_T \iota'_T \quad \text{and} \quad A = I_N \otimes A_T, \tag{18}$$

the LSDV estimator of δ, which is called the CV (covariance estimator) by Anderson and Hsiao (1981, p. 599), and is also often indicated as the fixed effects or the within-group estimator, can simply be expressed as

$$\hat{\delta} = (W'AW)^{-1} W'Ay. \tag{19}$$

This estimator is well-known to be inconsistent for finite T and $N \to \infty$, and so is the maximum likelihood estimator (see Anderson and Hsiao, 1982, p. 74). Although consistent estimators can be obtained by instrumental variables, we are nevertheless still interested in the inconsistent LSDV estimator. This is because it can be used to construct exact inference on γ, (see van den Doel and Kiviet, 1993), and also because it has a relatively small variance. Below, we shall derive an approximation to its small sample bias, and then show that a bias-corrected LSDV estimator can be constructed which compares favourably with other consistent ($N \to \infty$, fixed T) estimators.

Various implementations of IV (instrumental variable) estimators have been suggested which are consistent for either $T \to \infty$ or $N \to \infty$ or both. Anderson and Hsiao suggested the following procedure. Assuming that, next to y_{i0}, data are available for time period $t = 0$ for x_{it} too, we can define

$$\dot{y}_{it} = y_{it} - y_{i,t-1}, \quad \dot{x}_{it} = x_{it} - x_{i,t-1}, \qquad t = 1, \dots, T, \quad \forall i. \tag{20}$$

Assuming that data collection started at $t = 0$ and $T \geq 2$, we define

$$\dot{y}_i = \begin{bmatrix} \dot{y}_{i2} \\ \vdots \\ \dot{y}_{iT} \end{bmatrix}, \quad \dot{y}_i^{(-1)} = \begin{bmatrix} \dot{y}_{i1} \\ \vdots \\ \dot{y}_{iT-1} \end{bmatrix}, \quad \dot{X}_i = \begin{bmatrix} \dot{x}'_{i2} \\ \vdots \\ \dot{x}'_{iT} \end{bmatrix}, \quad \dot{\varepsilon}_i = \begin{bmatrix} \dot{\varepsilon}_{i2} \\ \vdots \\ \dot{\varepsilon}_{iT} \end{bmatrix}, \tag{21}$$

where $\dot{y}_i$, $\dot{y}_i^{(-1)}$, and $\dot{\varepsilon}_i$ are $(T-1) \times 1$ vectors and $\dot{X}_i$ is a $(T-1) \times K$ matrix. Upon stacking this information on all N individuals, we obtain the $N(T-1) \times 1$ vectors $\dot{y}$, $\dot{y}^{(-1)}$, and $\dot{\varepsilon}$ and the $N(T-1) \times K$ matrix $\dot{X}$, which are related according to

$$\dot{y} = \gamma \dot{y}^{(-1)} + \dot{X}\beta + \dot{\varepsilon} = \dot{W}\delta + \dot{\varepsilon} \quad \text{where} \quad \dot{W} = [\dot{y}^{(-1)} \; \vdots \; \dot{X}]. \tag{22}$$

A particular IV estimator suggested by Anderson and Hsiao uses instrumental variable set $Z_l = [y_l \; \vdots \; \dot{X}]$, where the first $N(T-1) \times 1$ instrumental vector is defined as

$$y_l = (y_{10}, \dots, y_{1,T-2}, y_{20}, \dots, y_{2,T-2}, \dots, y_{N0}, \dots, y_{N,T-2})'. \tag{23}$$

Below, we shall indicate this estimation procedure by $AH\mathscr{L}$. The same authors also suggest an IV estimator where the (twice) lagged first difference $\dot{y}^{(-2)}$ is used as an instrument instead of y_l (note that this can only be performed by dropping another observation in the time dimension, so it requires $T \geq 3$). We indicate this procedure by $AH\Delta$. However, Arellano (1989) showed that this estimator will suffer often from identification problems which will lead to a poor efficiency, so we will not pay much further attention to it.

Estimators which exploit more orthogonality conditions by employing more instrumental variables have been suggested *inter alia* by Holtz-Eakin et al. (1988), Arellano and Bond (1991), and Ahn and Schmidt (1993). Asymptotically the GMM (Generalized Method of Moments) estimators have optimal properties for particular classes of models. We shall only consider IV estimators which exploit particular linear moment restrictions. Following Arellano and Bond (1991) we examine in the Monte Carlo the procedures GMM1 and GMM2. Having $T + 1$ time-series observations available, these procedures use $K + T(T-1)/2$ instrumental variables for estimating (22); the two-step GMM2 estimator is robust with respect to correlated and heteroscedastic disturbances. In the context of our assumptions these estimators are not asymptotically optimal. However, this does not necessarily imply that their small sample behaviour will be inferior to the methods discussed in Ahn and Schmidt (1993).

Arellano and Bond find a small negative bias in the estimator for γ obtained by the GMM procedures. $AH\mathscr{L}$ shows a performance which is often similar to GMM, but can also be very imprecise, and even worse than $AH\Delta$. Arellano and Bond suppose this happens when γ is high; in Section 5 we come to somewhat different conclusions, which are less favourable for GMM.

When IV estimates are obtained from using as many instruments as regressors in the equation (just identification), such (indirect least-squares) estimators, as a rule, have no finite moments. In that light the unstable mean (squared) errors of both $AH\Delta$ and $AH\mathscr{L}$ are no surprise. Perhaps we can improve their performance by adding just one or a few instruments. Therefore, in the simulations in Section 5, we shall also examine results, indicated by $AHX\mathscr{L}$ and $AHX\Delta$, where apart from $\dot{X}$ and y_l or $\dot{y}^{(-2)}$ also the level of X is used as an instrument. As far as we know the actual existence of the moments of the estimators employed in this study has not yet been established. Any serious problems of this kind may emerge from the simulation results. For curiosity we shall also examine estimators where the untransformed original equation is estimated. Not only the inconsistent OLS estimator which neglects the individual effects will be calculated, but also IV estimators where we confront the original equation in levels with instruments which are in first differences or in deviation form (from individual means). From these instruments the individual effects have been removed, and therefore they are valid (except $Ay^{(-1)}$) and yield consistent estimators. We examine $IV\Delta X$ and $IVAX$. In the former, the instruments are first differences and conform to $\dot{W}$ given in (22). In the latter, the instruments $\dot{X}$ in $\dot{W}$ are replaced by AX.

3. The inconsistency of the LSDV estimator

The asymptotic bias of the LSDV estimator has been examined by Nickell (1981) in the fixed effects model, where he focusses on the case with random start-up (no conditioning on y_{i0}). For the model with no exogenous regressor variables he finds:

$$\operatorname*{plim}_{N\to\infty}(\hat{\gamma}-\gamma) = -\frac{1+\gamma}{T-1}\left[1-\frac{(1-\gamma^T)}{T(1-\gamma)}\right] \times\left\{1-\frac{2\gamma}{(1-\gamma)(T-1)}\left[1-\frac{(1-\gamma^T)}{T(1-\gamma)}\right]\right\}^{-1}, \tag{24}$$

which clearly shows that the inconsistency is $O(T^{-1})$ and negative for positive γ (if $T>1$); moreover, it does not depend on σ_ε^2. Even for small N it has been found to approximate the true bias, as assessed from Monte Carlo studies, pretty close, except for large values of γ. Similar results have been found by Sevestre and Trognon (1985). They consider the situation where the individual effects are random and examine the consequences of various assumptions regarding the initial observations. They do not consider just the LSDV estimator, but the class of λ-type estimators (see Maddala, 1971) which includes LSDV and OLS as special cases.

Earlier results on inconsistency in Trognon (1978) and Sevestre and Trognon (1983) also concern the case of dynamic panel data models with random exogenous individual effects, and include the case of a model with exogenous explanatory variables. Like Nickell (1981, p. 1424) they find that in this situation the asymptotic bias of $\hat{\gamma}$ is related to result (24) in such a way that it has to be multiplied now by

$$\left[\operatorname*{plim}_{N\to\infty} \frac{1}{N\cdot T} y^{(-1)'} A y^{(-1)}\right]\left[\operatorname*{plim}_{N\to\infty} \frac{1}{N\cdot T} y^{(-1)'} AMA y^{(-1)}\right]^{-1}, \tag{25}$$

where $M = I - AX(X'AX)^{-1}X'A$, whereas

$$\operatorname*{plim}_{N\to\infty} (\hat{\beta} - \beta) = -\left[\operatorname*{plim}_{N\to\infty}(X'AX)^{-1}X'Ay^{(-1)}\right]\cdot \operatorname*{plim}_{N\to\infty} (\hat{\gamma} - \gamma). \tag{26}$$

Nickell remarks that the factor (25) is greater than (or equal to) unity, so the presence of exogenous regressors aggravates the bias (although this is denied by Sevestre and Trognon).

Formulas (25) and (26) are not very helpful in providing a clearcut insight into the asymptotic bias, and they may even be very inaccurate as far as the actual magnitude of the bias of the LSDV estimator in small samples is concerned. Nickell (1981, p. 1423) gives an indication on how a more accurate approximation might be obtained. A comparable suggestion to approximate the bias to $O(N^{-1}T^{-1})$ is put forward by Beggs and Nerlove (1988), but they do not pursue this line of approach. Moreover, their suggestion seems only applicable for the approximation of the bias in $\hat{\gamma}$, and not for the complete coefficient vector $\hat{\delta}$. In the next section, however, we shall derive an approximating formula for the bias of the full vector of LSDV coefficient estimates in the model with exogenous regressors. As we shall show its magnitude can be evaluated and exploited easily.

4. An approximation to the bias of the LSDV estimator

Upon substituting (17), the bias of the LSDV estimator given in (19) can be expressed as (below expectations are always conditional on X and v_0):

$$\begin{aligned} \mathrm{E}(\hat{\delta} - \delta) &= \mathrm{E}(W'AW)^{-1}W'A[W\delta + (I_N \otimes \iota_T)\eta + \varepsilon] - \delta \\ &= \mathrm{E}(W'AW)^{-1}W'A\varepsilon. \end{aligned} \tag{27}$$

Here we made use of $A(I_N \otimes \iota_T) = (I_N \otimes A_T\iota_T) = 0$, since $A_T\iota_T = 0$. The expectation (27) is not easily evaluated since W is stochastic and hence the expression is nonlinear. In order to approximate this expectation it is helpful to expose the stochastic nature of W more explicitly by introducing a notation which clearly distinguishes between the fixed elements of W, viz. $\mathrm{E}(W)$, and the zero-mean random elements of W, i.e., $W - \mathrm{E}(W)$; we label them $\bar{W}$ and $\tilde{W}$, respectively. So, we introduce the following notational convention:

For any stochastic or nonstochastic $m_1 \times m_2$ matrix M we define $\bar{M} = \mathrm{E}(M)$ and $\tilde{M} = M - \mathrm{E}(M)$, and hence $M = \bar{M} + \tilde{M}$, where $\bar{M}$ is nonrandom and $\mathrm{E}(\tilde{M}) = 0$.

We thus have

$$W = \bar{W} + \tilde{W} \quad \text{where} \quad \bar{W} = [\bar{y}^{(-1)} \ \vdots \ X], \quad \tilde{W} = [\tilde{y}^{(-1)} \ \vdots \ 0]. \tag{28}$$

For further analysis of (27) we concentrate on AW, i.e., on $A\bar{W}$ and $A\tilde{W}$, and more in particular on $A\tilde{y}^{(-1)}$. We obtain

$$A\tilde{y}^{(-1)} = (I_N \otimes A_T) \begin{bmatrix} \tilde{y}_1^{(-1)} \\ \vdots \\ \tilde{y}_N^{(-1)} \end{bmatrix} = \begin{bmatrix} A_T\tilde{y}_1^{(-1)} \\ \vdots \\ A_T\tilde{y}_N^{(-1)} \end{bmatrix} = \begin{bmatrix} A_T\tilde{v}_1^{(-1)} \\ \vdots \\ A_T\tilde{v}_N^{(-1)} \end{bmatrix}, \tag{29}$$

where we made use of (15) and (9), which is expressed now as

$$y_i = v_i + \frac{1}{1-\gamma}\eta_i\iota_T. \tag{30}$$

Note that again the dummy variables (i.e., the within operation) removes the individual effects η_i, and hence the stochastic nature of AW originates fully from $\tilde{v} = (\tilde{v}_1, \ldots, \tilde{v}_N)'$. Since $\tilde{x}_{it} = 0$ and $\tilde{\varepsilon}_{it} = \varepsilon_{it}$, it follows straightforwardly from (10) that

$$\tilde{v}_{it} = \gamma\tilde{v}_{i,t-1} + \varepsilon_{it}, \qquad i = 1, \ldots, N, \quad t = 1, \ldots, T. \tag{31}$$

With $\tilde{v}_{i0} = 0$, since $\bar{v}_{i0} = v_{i0}$, relation (31) gives

$$\tilde{v}_i^{(-1)} = L\tilde{v}_i, \quad \Gamma\tilde{v}_i = \varepsilon_i \quad \text{and thus} \quad \tilde{v}_i^{(-1)} = L\Gamma^{-1}\varepsilon_i = C\varepsilon_i, \tag{32}$$

where L, Γ, and C are $T \times T$ matrices, viz.

$$L = \begin{bmatrix} 0 & 0 & \cdot & \cdot & \cdot & \cdot & 0 \\ 1 & 0 & & & & & \cdot \\ 0 & 1 & 0 & & & & \cdot \\ \cdot & \cdot & 1 & \cdot & & & \cdot \\ \cdot & & \cdot & \cdot & \cdot & & \cdot \\ \cdot & & & \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & \cdot & \cdot & 0 & 1 & 0 \end{bmatrix}, \qquad \Gamma = \begin{bmatrix} 1 & 0 & \cdot & \cdot & \cdot & \cdot & 0 \\ -\gamma & 1 & 0 & & & & \cdot \\ 0 & -\gamma & 1 & & & & \cdot \\ \cdot & \cdot & \cdot & \cdot & & & \cdot \\ \cdot & & \cdot & \cdot & \cdot & & \cdot \\ \cdot & & & \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & \cdot & \cdot & 0 & -\gamma & 1 \end{bmatrix},$$

$$C = \begin{bmatrix} 0 & \cdot & \cdot & \cdot & \cdot & \cdot & 0 \\ 1 & 0 & & & & & \cdot \\ \gamma & 1 & 0 & & & & \cdot \\ \gamma^2 & \gamma & 1 & \cdot & & & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & & \cdot \\ \cdot & & & \cdot & \cdot & \cdot & \cdot \\ \gamma^{T-2} & \cdot & \cdot & \cdot & \gamma & 1 & 0 \end{bmatrix}.$$

Hence, from (29) and (35) we find

$$A\tilde{y}^{(-1)} = A\tilde{v}^{(-1)} = (I_N \otimes A_T)(I_N \otimes C)\varepsilon = (I_N \otimes A_T C)\varepsilon. \tag{33}$$

So, according to (28), we obtain for $A\tilde{W}$ (the stochastic elements of AW)

$$A\tilde{W} = (I_N \otimes A_T C)\varepsilon q', \tag{34}$$

where q is the $(K+1) \times 1$ unit vector

$$q = (1, 0, \dots, 0)'. \tag{35}$$

Note that (34) shows that the stochastic characteristics of AW are linear in the disturbance vector ε.

Formula (34) enables to find an approximation to the bias (27) when use is made of the following lemma (for proofs see Appendix A).

Lemma. Let D be a $m \times m$ stochastic matrix, such that $\bar{D} = \mathrm{E}(D)$ exists with $\bar{D}$ nonsingular, and

$$D = \sum_{j=1}^{n} d_j d_j',$$

where the d_j are $m \times 1$ vectors such that

$$\bar{D} = \mathrm{O}(n), \qquad D = \mathrm{O}_p(n), \qquad (D - \bar{D}) = \mathrm{O}_p(n^{1/2}),$$

then we have

$$D^{-1} = (\bar{D})^{-1} - (\bar{D})^{-1}(D - \bar{D})(\bar{D})^{-1} + \mathrm{O}_p(n^{-2}).$$

We also make use of the well-known result that for general nonstochastic $m \times m$ matrices Q_1 and Q_2 and a random m-element vector $\xi \sim \mathrm{N}(0, \sigma_\varepsilon^2 I_m)$ we have

$$\mathrm{E}(\xi' Q_1 \xi)(\xi' Q_2 \xi) = \sigma_\varepsilon^4 [\mathrm{tr}(Q_1) \cdot \mathrm{tr}(Q_2) + \mathrm{tr}(Q_1 Q_2 + Q_1' Q_2)]. \tag{36}$$

For ease of reference, we now state the following:

Assumption 1. Let (1) through (7) hold, i.e., model (17) is a valid specification of the relationship between observations y and X (including the assumption that the regressors are stationary).

In Appendix A we prove:

Theorem 1. Under Assumption 1 we find for the LSDV estimator $\hat{\delta}$ of (19):

$$\begin{aligned}
\mathrm{E}(\hat{\delta} - \delta) = -\sigma_\varepsilon^2 (\bar{D})^{-1} \Bigg(& \frac{N}{T}(\iota_T' C \iota_T)[2q - \bar{W}' A \bar{W} (\bar{D})^{-1} q] \\
& + \mathrm{tr}\{\bar{W}'(I_N \otimes A_T C A_T) \bar{W} (\bar{D})^{-1}\} q \\
& + \bar{W}'(I_N \otimes A_T C A_T) \bar{W} (\bar{D})^{-1} q + \sigma_\varepsilon^2 N q' (\bar{D})^{-1} q \\
& \times \left[-\frac{N}{T}(\iota_T' C \iota_T) \mathrm{tr}\{C' A_T C\} + 2 \mathrm{tr}\{C' A_T C A_T C\} \right] q \Bigg) \\
& + \mathrm{O}(N^{-1} T^{-3/2}),
\end{aligned}$$

where $\bar{D} = \bar{W}' A \bar{W} + \sigma_\varepsilon^2 N \mathrm{tr}\{C' A_T C\} q q'$ and $A\bar{W} = \mathrm{E}(AW)$, with W, A, A_T, ι_T, q, and C given in (16), (18), (12), (35), and *(32) respectively.*

The approximation consists of terms which are either $\mathrm{O}(1/T)$ or $\mathrm{O}(1/NT)$.

We also want to cover the case where assumption (7) is relaxed, and where we allow regressors to be either stationary or nonstationary. To that end we can formulate:

Assumption 2. Let (1) through (6) hold, and let $\Lambda = (\lambda_{ii})$ be a $(K+1) \times (K+1)$ diagonal matrix of normalizing factors, with $\lambda_{ii} = (T^{-1/2})^{d(i)}$ and the integer $d(i) \geqslant 0$ $(i = 1, \ldots, K+1)$, such that for $B = \Lambda W' A W \Lambda$ we have $B = \mathrm{O}_P(NT)$, $\bar{B} = \mathrm{O}(NT)$, and $(B - \bar{B}) = \mathrm{O}_p(N^{1/2} T^{1/2})$.

Obviously, the diagonal elements of Λ which correspond with stationary regressors are taken unity, and the others are typically $T^{-1/2}$ or T^{-1} etc. We have:

Theorem 2. Under Assumption 2 the bias of the LSDV estimator $\hat{\delta}$ of (19) can be approximated by the same first term as given in Theorem 1, but now the remainder term for the ith element ($i = 1, \ldots, K + 1$) of the coefficient estimate $\hat{\delta}_i$ is $O(N^{-1}T^{-3/2-d(i)})$.

The approximation obtained in Theorems 1 and 2 is used below in two ways. Firstly, we perform simulations in order to assess the actual small sample bias of the LSDV estimator, and then check the accuracy of the approximative formula. The latter can only be evaluated in a situation where the true parameters δ and σ_ε^2 and v_{i0} are known, as is the case in a simulation study. Secondly, however, we also examine in the simulation study the performance of an operational corrected LSDV estimation procedure, labelled LSDVc. Here, an estimate of the approximation formula is subtracted from the original LSDV estimator. For this estimate we will use the GMM1 results. This new estimation procedure will be compared with GMM1 and the various other estimation methods mentioned in Section 2. We assess the bias, the true standard error, and the root mean-squared error of the various estimators of the coefficients γ and β from 1000 Monte Carlo experiments for a few relevant designs.

5. Simulation experiments

In our simulation design we have to make a balanced and relevant choice on the various characteristics of model (1). From the result in Theorem 1 it is obvious that N, T, σ_ε^2, and γ (which appears in C) are important. The matrix X and its coefficients β affect the bias of the LSDV esimator through $A\bar{W} = [A\bar{y}^{(-1)} : AX]$; its first column can be established from (30) which yields $A\bar{y}_i = A\bar{v}_i$, where $\bar{v}_{it} = \gamma\bar{v}_{i,t-1} + x_{it}'\beta$. Obviously v_{i0} is important. Some estimators, viz. $LSDV$, $AH\Delta$, and $AHX\Delta$, are invariant with respect to η (and hence to σ_η^2); others are not. We perform the Monte Carlo simulation study on a design that closely corresponds to the experiments presented in Arellano and Bond (1991)–AB henceforth. However, our design is different in four respects.

Firstly, we do not fix $y_{i0} = 0$ while discarding the first ten cross-sections. We generate the data exactly according to the specifications, avoiding any slow convergence problems and waste of random numbers (see Appendix B).

Like AB we shall use σ_ε^2 as the normalization parameter, i.e., $\sigma_\varepsilon^2 = 1$, but a second difference with AB is that we do not siply choose $\sigma_\eta^2 = \sigma_\varepsilon^2$, but

$$\sigma_\eta = \mu\sigma_\varepsilon(1 - \gamma), \qquad \mu \geqslant 0. \tag{37}$$

The reason for including the factor $(1 - \gamma)$ follows from (9). By choosing $\mu = 1$ the impact on y_{it} of the disturbance and the individual effect have equal magnitudes. For $\mu = 2$ the impact of the effect is double (irrespective of any changes in γ). In AB changes in the dynamics through γ also affect the impact ratio of the two error components ε_{it} and η_i, which then obscures to establish the separate effects of γ and μ.

Like AB we include only one regressor variable, but a third difference is that we do not fix $\beta = 1$, but vary β (like σ_η) with γ such that the long-run multiplier of x with respect to y is unity, hence we choose

$$\beta = 1 - \gamma. \tag{38}$$

Now a change in γ only changes the dynamics of the relationship, and not at the same time the equilibrium causal relationship between y and x. The multiplier of unity does not lead to any loss of generality, since the contribution of βx_{it} in the relationship can also be varied by changing the scale of the regressor variable.

The generating equation for the exogenous explanatory variable is

$$x_{it} = \rho x_{i,t-1} + \xi_{it} \quad \text{with} \quad \xi_{it} \sim \text{i.i.d. N}(0, \sigma_\xi^2). \tag{39}$$

Therefore, apart from the coefficients β (which is linked to γ here) and ρ, it is also σ_ξ^2 which determines the signal-to-noise ratio of the relationship. AB give little or no attention to this ratio, although it will be of great significance for the magnitude of bias and the general performance of the various estimation techniques. Even if the only concern is the relative bias and efficiency of the estimators under investigation then, through nonlinear dependencies, the signal-to-noise ratio may still be of considerable importance. Therefore, we shall separately choose values for σ_ε^2, σ_η^2 (via μ), and the signal σ_s^2, which we define as the variance of the right-hand side of (10) minus the (independent) contribution of the disturbance. Upon using (B.3) and (B.7) we find:

$$\begin{aligned}\sigma_s^2 &= \text{var}(v_{it} - \varepsilon_{it}) \\ &= \beta^2 \sigma_\xi^2 \left[1 + \frac{(\gamma + \rho)^2}{1 + \gamma\rho}[\gamma\rho - 1] - (\gamma\rho)^2\right]^{-1} + \frac{\gamma^2}{1 - \gamma^2}\sigma_\varepsilon^2. \end{aligned} \tag{40}$$

When varying γ (and thus β) in the simulations, we shall also adapt the value of σ_ξ^2 in such a way that we control the signal. It follows from (40) that a signal with variance σ_s^2 is obtained by taking

$$\sigma_\xi^2 = \left[\sigma_s^2 - \frac{\gamma^2}{1 - \gamma^2}\sigma_\varepsilon^2\right]\left[1 + \frac{(\gamma + \rho)^2}{1 + \gamma\rho}[\gamma\rho - 1] - (\gamma\rho)^2\right]\frac{1}{\beta^2}. \tag{41}$$

This choice is the fourth difference from the AB design, who have $\sigma_\xi^2 = 0.9$. From the first factor in square brackets of (41) we see that not all values of γ, σ_ε^2, and σ_s^2 are compatible (lead to positive σ_ξ^2).

By (41) the signal will be controlled only to a limited degree since the generation of the x_{it} is not replicated in every simulation run; they are generated only once (like in AB), and hence, in this design the exogenous variable x_{it} is independent of the individual effects η_i. Many of the techniques investigated here, however, are designed for the situation where the observed and the

Table 1
Monte Carlo design, 14 different parameter combinations

	T	γ	β	ρ	σ_ς^2	μ	σ_η	σ_ξ	σ_X	R^2
I	6	0.0	1.0	0.8	2	1	1.0	0.85	1.41	0.78
II	6	0.4	0.6	0.8	2	1	0.6	0.88	1.47	0.79
III	6	0.8	0.2	0.8	2	1	0.2	0.40	0.66	0.81
IV	6	0.0	1.0	0.99	2	1	1.0	0.20	1.41	0.62
V	6	0.4	0.6	0.99	2	1	0.6	0.19	1.35	0.66
VI	6	0.8	0.2	0.99	2	1	0.2	0.07	0.48	0.81
VII	3	0.4	0.6	0.8	2	1	0.6	0.88	1.47	0.84
VIII	3	0.4	0.6	0.8	8	1	0.6	1.84	3.06	0.97
IX	3	0.4	0.6	0.8	2	5	3.0	0.88	1.47	0.98
X	3	0.4	0.6	0.8	8	5	3.0	1.84	3.06	0.99
XI	3	0.4	0.6	0.99	2	1	0.6	0.19	1.35	0.76
XII	3	0.4	0.6	0.99	8	1	0.6	0.40	2.81	0.83
XIII	3	0.4	0.6	0.99	2	5	3.0	0.19	1.35	0.98
XIV	3	0.4	0.6	0.99	8	5	3.0	0.40	2.81	0.98

unobserved individual heterogeneity are correlated. Note that *LSDV* (and also *AHΔ*) are invariant with respect to η. The effects on other techniques may be substantial, but will not be examined here.

To sum up, we will pay special attention to the parameter values:

$$\gamma = 0.0,\ 0.4,\ 0.8, \tag{42}$$

$$\rho = 0.8,\ 0.99,$$

$$\mu = 1,\ 5,$$

$$\sigma_\varepsilon^2 = 1, \qquad \sigma_\varsigma^2 = 2,\ 8.$$

The latter values are feasible and may differ sharply from those in the design of AB. For the design parameter values of their table 1 ($\gamma = 0.2, 0.5, 0.8$, $\rho = 0.8$, $\beta = 1$, $\sigma_\xi^2 = 0.9$), signal variances are found of 3.6, 8.1, 33.4, with $\mu = 1.25, 2, 5$, respectively. Hence, slower adjustment is conflated with more substantial signal and individual effects.

We investigate samples of size:

$$N = 100 \quad \text{and} \quad T = 3,\ 6. \tag{43}$$

(AB have $N = 100$, $T = 7$, but their T denotes the total number of observations available.) Choices (42) and (43) yield 48 combinations. Here we present results on only 14 of these (a copy of all computer output is available upon request). Table 1 presents the design parameter value combinations of these 14 selected

Table 2
Monte Carlo results for $\rho = 0.8$, $\sigma_s^2 = 2$, $\mu = 1$, $T = 6$

		Bias γ	Bias β	Std γ	Std β	Rmse γ	Rmse β
I	*GMM1*	– 0.036	– 0.015	0.058	0.070	0.068	0.071
	GMM2	– 0.033	– 0.026	0.065	0.074	0.073	0.078
	AHΔ	0.002	– 0.004	0.117	0.083	0.117	0.083
	AHXΔ	0.009	– 0.003	0.101	0.083	0.102	0.083
	AH$\mathscr{L}$	– 0.021	– 0.009	0.064	0.073	0.067	0.073
	AHX$\mathscr{L}$	0.218	0.287	0.094	0.104	0.238	0.305
	IVAX	0.001	0.005	0.061	0.067	0.061	0.067
	IVΔX	– 0.009	0.060	0.052	0.131	0.053	0.144
	OLS	0.370	– 0.345	0.046	0.055	0.373	0.349
	OLSι	0.368	– 0.342	0.046	0.055	0.371	0.346
	LSDV	– 0.111	0.020	0.035	0.054	0.117	0.058
	LSDVb	– 0.004	– 0.023	0.034	0.054	0.035	0.059
	LSDVc	– 0.019	– 0.018	0.038	0.054	0.043	0.057
II	*GMM1*	– 0.050	– 0.002	0.079	0.067	0.093	0.067
	GMM2	– 0.045	– 0.011	0.088	0.071	0.099	0.072
	AHΔ	0.022	0.000	0.273	0.082	0.274	0.082
	AHXΔ	0.013	0.001	0.166	0.079	0.166	0.079
	AH$\mathscr{L}$	– 0.018	– 0.004	0.092	0.070	0.094	0.070
	AHX$\mathscr{L}$	0.599	0.184	0.187	0.115	0.627	0.216
	IVAX	– 0.000	0.003	0.072	0.058	0.072	0.058
	IVΔX	– 0.007	0.026	0.061	0.102	0.061	0.105
	OLS	0.204	– 0.171	0.033	0.039	0.207	0.176
	OLSι	0.203	– 0.169	0.033	0.039	0.206	0.174
	LSDV	– 0.187	0.039	0.039	0.052	0.191	0.065
	LSDVb	– 0.003	– 0.008	0.039	0.052	0.039	0.052
	LSDVc	– 0.038	– 0.002	0.045	0.052	0.059	0.052
III	*GMM1*	– 0.065	0.000	0.099	0.155	0.118	0.155
	GMM2	– 0.065	– 0.006	0.112	0.166	0.129	0.166
	AHΔ	0.235	0.023	4.706	0.819	4.712	0.819
	AHXΔ	0.037	0.007	0.600	0.193	0.601	0.193
	AH$\mathscr{L}$	– 0.002	– 0.001	0.131	0.159	0.131	0.159
	AHX$\mathscr{L}$	2.109	0.176	1.891	0.521	2.832	0.550
	IVAX	– 0.009	0.006	0.085	0.119	0.086	0.119
	IVΔX	– 0.012	0.053	0.085	0.230	0.085	0.236
	OLS	0.049	– 0.068	0.021	0.058	0.054	0.090
	OLSι	0.048	– 0.067	0.021	0.058	0.053	0.089
	LSDV	– 0.360	0.005	0.042	0.117	0.362	0.117
	LSDVb	– 0.005	0.007	0.042	0.113	0.042	0.113
	LSDVc	– 0.125	– 0.011	0.049	0.113	0.135	0.114

Table 3
Monte Carlo results for $\rho = 0.99$, $\sigma_s^2 = 2$, $\mu = 1$, $T = 6$

		Bias γ	Bias β	Std γ	Std β	Rmse γ	Rmse β
IV	*GMM1*	− 0.018	0.004	0.062	0.311	0.065	0.311
	GMM2	− 0.018	− 0.016	0.070	0.335	0.073	0.335
	$AH\Delta$	0.002	0.003	0.115	0.352	0.115	0.352
	$AHX\Delta$	0.003	0.003	0.115	0.352	0.115	0.352
	$AH\mathscr{L}$	− 0.004	0.006	0.069	0.318	0.069	0.318
	$AHX\mathscr{L}$	0.015	0.277	0.088	0.407	0.089	0.492
	IVAX	− 0.009	0.083	0.076	0.234	0.076	0.248
	$IV\Delta X$	− 0.036	0.982	0.102	0.739	0.108	1.229
	OLS	0.498	− 0.872	0.047	0.040	0.500	0.873
	OLSι	0.495	− 0.870	0.048	0.041	0.498	0.871
	LSDV	− 0.164	0.087	0.039	0.226	0.168	0.242
	LSDVb	− 0.029	0.022	0.039	0.224	0.049	0.225
	LSDVc	− 0.017	0.005	0.047	0.217	0.050	0.217
V	*GMM1*	− 0.033	0.009	0.078	0.326	0.084	0.326
	GMM2	− 0.030	− 0.005	0.088	0.351	0.093	0.351
	$AH\Delta$	0.013	0.006	0.220	0.371	0.220	0.371
	$AHX\Delta$	0.014	0.006	0.216	0.370	0.216	0.370
	$AH\mathscr{L}$	− 0.004	0.009	0.091	0.333	0.091	0.333
	$AHX\mathscr{L}$	0.466	0.176	0.285	0.505	0.546	0.535
	IVAX	− 0.011	0.054	0.084	0.222	0.084	0.229
	$IV\Delta X$	− 0.029	0.601	0.106	0.634	0.110	0.874
	OLS	0.271	− 0.517	0.034	0.032	0.273	0.518
	OLSι	0.270	− 0.516	0.034	0.033	0.272	0.517
	LSDV	− 0.249	0.104	0.042	0.242	0.253	0.263
	LSDVb	− 0.040	0.049	0.042	0.238	0.057	0.243
	LSDVc	− 0.042	0.019	0.052	0.230	0.067	0.231
VI	*GMM1*	− 0.062	0.031	0.098	0.914	0.115	0.914
	GMM2	− 0.061	0.014	0.110	0.984	0.126	0.984
	$AH\Delta$	0.621	0.101	12.032	7.892	12.048	7.892
	$AHX\Delta$	0.045	− 0.009	1.014	1.297	1.015	1.297
	$AH\mathscr{L}$	0.000	0.034	0.129	0.941	0.129	0.942
	$AHX\mathscr{L}$	3.055	− 0.021	4.653	4.975	5.567	4.975
	IVAX	− 0.008	0.039	0.088	0.579	0.089	0.580
	$IV\Delta X$	− 0.008	0.262	0.102	1.574	0.103	1.595
	OLS	0.050	− 0.180	0.021	0.082	0.054	0.198
	OLSι	0.049	− 0.180	0.021	0.084	0.053	0.199
	LSDV	− 0.365	0.089	0.042	0.668	0.367	0.674
	LSDVb	− 0.008	0.100	0.041	0.643	0.042	0.650
	LSDVc	− 0.127	0.050	0.049	0.641	0.136	0.643

experiments. It also gives the resulting values of σ_η, σ_ξ, $\sigma_x = [\mathrm{var}(x_{it})]^{1/2}$; the R^2 column contains the simulation average of this coefficient as obtained in the LSDV regressions. Note that LSDV produces biased and inconsistent estimates; nevertheless the R^2 values give some indication of the fit. The number of simulation replications was set at 1000.

Tables 2, 3, and 4 present the bias, the standard error (std), and the root mean squared error (rmse) of the various estimation techniques for both γ and β. Table 2 gives results for design parameter values $\rho = 0.8$, $\sigma_s^2 = 2$, $\mu = 1$ at $T = 6$. We see that GMM is virtually unbiased as far as β is concerned, but the γ estimate has a negative bias which is over 50% of its standard deviation. $AH\Delta$ is much less biased in designs I and II, at the expense of a loss in efficiency, but it breaks down for $\gamma = 0.8$, whereas $AHX\Delta$ seems more stable, as far as the bias is concerned, but the efficiency loss at $\gamma = 0.8$ is substantial. $AH\mathscr{L}$ shows smaller biases than GMM and its efficiency compares quite favourably. $AHX\mathscr{L}$, however, seems a waster. $IVAX$ is found to have remarkable capabilities. In these designs it beats GMM. It is virtually unbiased, and its efficiency is favourable, especially for the higher γ value. $IV\Delta X$ proves to have problems with estimating β, but as far as γ is concerned it beats GMM as well. It has already been established in many earlier studies, that OLS estimates γ with positive bias in this situation. Moreover, we find here that it has impressingly small standard deviation, and therefore, when the bias is moderate (which it is for the higher γ value) it has an attractive mean squared error and, in this situation, GMM is beaten again. The results are more or less the same when an intercept is included in the regressions ($OLS\imath$).

We now turn to the results on LSDV and its bias corrected variants. That the LSDV γ estimator has negative (asymptotic) bias is well-known from the literature. We see that this bias gets extremely large when γ increases (whereas the bias in β remains moderate). Hence, for slower dynamic adjustment processes, the bias is more serious (this regularity could not be established from the Arellano and Bond design). Although the standard deviation of the LSDV estimators is not as attractive as the OLS dispersion, for γ it is smaller than achieved by any of the consistent estimators. By LSDVb we indicate results which are obtained by subtracting from the LSDV estimator the bias approximation formula evaluated for the true values of the parameters and $\bar{W}$. Since LSDVb shows hardly any bias, we learn that at these parameter values the approximation is amazingly accurate. Since v_{i0} differs in each replication, the bias approximation varies for each replication, and so both estimators (LSDV and LSDVb) do not necessarily have equivalent standard deviations. LSDVb has the highest efficiency of all estimators considered here, but, as we know, it is not an operational technique. An operational variant is LSDVc, where GMM1 results are used to assess the bias formula. Obviously, whenever GMM1 shows noticable bias, the bias correction will be less accurate, and therefore LSDVc can be much more biased than LSDVb, and also than, for instance, $AH\mathscr{L}$. However,

Table 4
Monte Carlo results for $\gamma = 0.4$, $T = 3$

		Bias γ	Bias β	Std γ	Std β	Rmse γ	Rmse β
VII	*GMM1*	– 0.049	– 0.006	0.181	0.108	0.188	0.109
	LSDV	– 0.395	0.012	0.061	0.091	0.400	0.092
	LSDVb	– 0.010	0.006	0.058	0.088	0.059	0.088
	LSDVc	– 0.205	0.004	0.081	0.093	0.220	0.093
VIII	*GMM1*	0.255	0.073	0.147	0.062	0.295	0.096
	LSDV	– 0.175	0.051	0.036	0.045	0.178	0.068
	LSDVb	– 0.029	0.029	0.036	0.044	0.046	0.053
	LSDVc	– 0.033	0.061	0.069	0.048	0.077	0.077
IX	*GMM1*	– 0.096	– 0.008	0.389	0.112	0.401	0.112
	LSDV	– 0.395	0.012	0.061	0.091	0.400	0.092
	LSDVb	– 0.010	0.006	0.058	0.088	0.059	0.088
	LSDVc	– 0.211	0.007	0.105	0.094	0.236	0.094
X	*GMM1*	0.175	0.063	0.249	0.067	0.305	0.092
	LSDV	– 0.175	0.051	0.036	0.045	0.178	0.068
	LSDVb	– 0.029	0.029	0.036	0.044	0.046	0.053
	LSDVc	– 0.052	0.059	0.082	0.048	0.097	0.076
XI	*GMM1*	– 0.026	0.005	0.167	0.506	0.169	0.506
	LSDV	– 0.495	0.101	0.066	0.430	0.499	0.442
	LSDVb	– 0.086	0.070	0.065	0.409	0.108	0.415
	LSDVc	– 0.249	0.055	0.083	0.430	0.263	0.433
XII	*GMM1*	– 0.041	0.003	0.192	0.243	0.196	0.243
	LSDV	– 0.484	0.087	0.065	0.207	0.488	0.224
	LSDVb	– 0.159	0.057	0.065	0.198	0.172	0.206
	LSDVc	– 0.248	0.046	0.087	0.207	0.263	0.212
XIII	*GMM1*	– 0.060	0.008	0.341	0.512	0.346	0.512
	LSDV	– 0.495	0.101	0.066	0.430	0.499	0.442
	LSDVb	– 0.086	0.070	0.065	0.409	0.108	0.415
	LSDVc	– 0.263	0.061	0.102	0.432	0.282	0.437
XIV	*GMM1*	– 0.064	0.005	0.349	0.246	0.355	0.246
	LSDV	– 0.484	0.087	0.065	0.207	0.488	0.224
	LSDVb	– 0.159	0.057	0.065	0.198	0.172	0.206
	LSDVc	– 0.258	0.052	0.103	0.208	0.278	0.215

since its standard deviation is still close to the original favourable LSDV dispersion, LSDVc has a mean squared error which, although not uniformly superior, can compete with any other of the operational techniques. It seems worthwhile to investigate other variants of LSDVc, where we do not stick to the GMM results for the bias assessment.

Table 3 contains the same experiments with the only difference that the exogenous regressor is extremely smooth now (although still stationary). At particular coefficient values $AH\Delta$, $AHX\mathscr{L}$, and $IV\Delta X$ break down. $AH\mathscr{L}$ compares well with GMM, and $IVAX$ is even better. Oddly enough OLS works quite well in design VI. LSDVc is usually better than GMM and than most other techniques.

On the whole, the Anderson-Hsiao and the IV estimates of the equation in levels are found to be very vulnerable with respect to the parameter values. In the final Table 4 we focus on GMM1 and LSDV correction. We find that even for very small T the bias approximation is often very accurate and the efficiency of LSDVb is always superior; on balance LSDVc seems more attractive than GMM1.

6. Conclusions

From this study it is found that for dynamic panel data models the use of instrumental variables estimation methods may lead to poor finite sample efficiency. In particular situations it seems that valid orthogonality restrictions can better not be employed when composing a set of instrumental variables. It is difficult to find clues on when which instrumental variables are better put aside in order to avoid serious small sample bias or relatively large standard deviations, which both entail poor estimator efficiency. As yet, no technique is available that has shown uniform superiority in finite samples over a wide range of relevant situations as far as the true parameter values and the further properties of the data generating mechanism are concerned. Perhaps such a technique is just impossible. However, from our Monte Carlo experiments it follows that in many circumstances a bias corrected version of the (inconsistent) LSDV estimator is surprisingly efficient in comparison to the established consistent estimation methods, whereas this property seems rather robust. This is a result of the fact that the LSDV estimator usually has a much smaller dispersion than IV estimators, whilst the bias correction of this estimator hardly enhances its variance.

The assessment of the bias via the approximation formula derived in this paper can be extremely accurate, despite the mostly very substantial bias in LSDV. Frequently, however, there seems to be room for some further improvements (occasionally the bias correction has actually produced a larger bias). Such improvements might be achieved by either using a different (perhaps iterated) initial estimate of the parameters than one obtained from the sometimes pretty inefficient GMM estimator, or by further analytical work which would yield a more accurate approximation. The concomitant errors of the present approach are $O(N^{-1}T^{-3/2})$, hence the corrected estimator (LSDVc) is now consistent not just for $T \to \infty$, but also for finite T and $N \to \infty$. However, substantial further analytical derivations may reduce the errors to $o(N^{-1}T^{-1})$ which then would yield corrected estimators which require also a less substantial value of N.

The results in Kiviet and Phillips (1994) show that a generalization for the model with higher-order lags is feasible. At this stage the major conclusion about the present bias approximation formula is that the corrected LSDV estimator is reasonably easy to calculate and was never outclassed in our simulations by any of the other estimators, whereas most of these standard IV or GMM estimation techniques have been found to produce very poor results in particular customary finite sample situations. Whether this will also be the case for asymptotically efficient GMM estimators and in situations where regressors and effects are correlated is still an unanswered question.

Appendix A

Proof of the Lemma. See Kiviet and Phillips (1994, Lemma 2).

Proof of Theorem 1. In the derivation of the approximation we take $D = W'AW$, when applying the Lemma. We first obtain some results on this matrix in order to check whether the conditions of the lemma are satisfied. We find

$$\begin{aligned} \mathrm{E}(W'AW) &= \mathrm{E}(\bar{W}'A\bar{W} + \tilde{W}'A\bar{W} + \bar{W}'A\tilde{W} + \tilde{W}'A\tilde{W}) \\ &= \bar{W}'A\bar{W} + \mathrm{E}(\tilde{W}'A\tilde{W}), \end{aligned} \tag{A.1}$$

and using (34),

$$\begin{aligned} \mathrm{E}(\tilde{W}'A\tilde{W}) &= \mathrm{E}q\varepsilon'(I_N \otimes C'A_T)(I_N \otimes A_T C)\varepsilon q' \\ &= qq'\mathrm{E}\varepsilon'(I_N \otimes C'A_T)(I_N \otimes A_T C)\varepsilon \\ &= \sigma_\varepsilon^2 \operatorname{tr}\{I_N \otimes C'A_T C\} qq' \\ &= \sigma^2 N \operatorname{tr}\{C'A_T C\} qq' \\ &= O(NT). \end{aligned} \tag{A.2}$$

Moreover, employing (36) we find

$$\begin{aligned} \operatorname{var}[\varepsilon'(I_N \otimes C'A_T)(I_N \otimes A_T C)\varepsilon] &= \mathrm{E}\varepsilon'(I_N \otimes C'A_T C)\varepsilon\varepsilon'(I_N \otimes C'A_T C)\varepsilon \\ &\quad - [\sigma_\varepsilon^2 N \operatorname{tr}\{C'A_T C\}]^2 \\ &= 2\sigma_\varepsilon^4 \operatorname{tr}\{(I_N \otimes C'A_T CC'A_T C)\} \\ &= O(NT), \end{aligned}$$

giving, with (A.2),

$$\tilde{W}'A\tilde{W} - \mathrm{E}(\tilde{W}'A\tilde{W}) = \mathrm{O}_p(N^{1/2}T^{1/2}). \tag{A.3}$$

Because of (5) and (7) all regressors, including the lagged dependent variable, are assumed to be stationary, hence:

$$W'AW = \mathrm{O}_p(NT) \quad \text{and} \quad \mathrm{E}(W'AW) = \mathrm{O}(NT). \tag{A.4}$$

We also have

$$\bar{W}'A\tilde{W} = \mathrm{O}_p(N^{1/2}T^{1/2}), \tag{A.5}$$

which follows from $\mathrm{E}(\bar{W}'A\tilde{W}) = 0$, and

$$\begin{aligned} \mathrm{var}(\bar{W}'A\tilde{y}^{(-1)}) &= \sigma_\varepsilon^2 \bar{W}'A(I_N \otimes CC')A\bar{W} \\ &= \mathrm{O}(NT). \end{aligned} \tag{A.6}$$

Hence, from (A.3) and (A.5), we find

$$\begin{aligned} D - \bar{D} &= \tilde{W}'A\bar{W} + \bar{W}'A\tilde{W} + [\tilde{W}'A\tilde{W} - \mathrm{E}(\tilde{W}'A\tilde{W})] \\ &= \mathrm{O}_p(N^{1/2}T^{1/2}), \end{aligned} \tag{A.7}$$

and from (A.4) and (A.7) it follows that we can invoke the lemma for $(W'AW)^{-1}$ in the bias expression, with $n = NT$.

Before we do so, we first consider the remaining part of the bias expression, viz. $W'A\varepsilon$. It is easily found now that

$$\begin{aligned} \mathrm{E}(W'A\varepsilon) &= \mathrm{E}(\bar{W}'A\varepsilon + \tilde{W}'A\varepsilon) \\ &= 0 + q\mathrm{E}\varepsilon'(I_N \otimes C'A_T)\varepsilon \\ &= \sigma_\varepsilon^2 N \cdot \mathrm{tr}\{A_T C\} q \\ &= -\sigma^2(\iota_T' C \iota_T)\frac{N}{T}q \\ &= \mathrm{O}(N), \end{aligned} \tag{A.8}$$

and

$$\begin{aligned} \mathrm{var}(W'A\varepsilon) &= \mathrm{E}(\bar{W}'A\varepsilon + \tilde{W}'A\varepsilon)(\varepsilon'A\bar{W} + \varepsilon'A\tilde{W}) - [\sigma_\varepsilon^2 N \cdot \mathrm{tr}\{A_T C\}]^2 qq' \\ &= \sigma_\varepsilon^2 \bar{W}'A\bar{W} + \mathrm{E}(\bar{W}'A\varepsilon\varepsilon'A\tilde{W} + \tilde{W}'A\varepsilon\varepsilon'A\bar{W} + \tilde{W}'A\varepsilon\varepsilon'A\tilde{W}) \\ &\quad - [\sigma_\varepsilon^2 N \cdot \mathrm{tr}\{A_T C\}]^2 qq'. \end{aligned}$$

Of these five terms, the second and third terms are zero, since they involve the third moment of normal variables. Thus,

$$\begin{aligned}\operatorname{var}(W'A\varepsilon) &= \sigma_\varepsilon^2 \bar{W}'A\bar{W} + qq'\{\mathrm{E}\varepsilon'(I_N \otimes C'A_T)\varepsilon\varepsilon'(I_N \otimes A_T C)\varepsilon \\ &\quad - [\sigma_\varepsilon^2 N \cdot \operatorname{tr}\{A_T C\}]^2\} \\ &= \sigma_\varepsilon^2 \bar{W}'A\bar{W} + qq'\sigma_\varepsilon^4 N[\operatorname{tr}\{C'A_T C\} + \operatorname{tr}\{A_T C A_T C\}] \\ &= \mathrm{O}(NT), \end{aligned} \tag{A.9}$$

and it follows that

$$W'A\varepsilon - \mathrm{E}(W'A\varepsilon) = \mathrm{O}_p(N^{1/2} \cdot T^{1/2}). \tag{A.10}$$

Finally we turn to the expression for the estimation error

$$\hat{\delta} - \delta = (W'AW)^{-1}\mathrm{E}(W'A\varepsilon) + (W'AW)^{-1}[W'A\varepsilon - \mathrm{E}(W'A\varepsilon)]. \tag{A.11}$$

When writing $D = W'AW$ and applying the lemma, the approximation of $(D)^{-1}$ by $[(\bar{D})^{-1} - (\bar{D})^{-1}(D - \bar{D})(\bar{D})^{-1}]$ leads to the omission of terms of $\mathrm{O}_p(N^{-2}T^{-2})$. Hence, by replacing $(D)^{-1}$ in the two terms of (A.11) we have an approximation of $(\hat{\delta} - \delta)$, where terms are omitted which are $\mathrm{O}_p(N^{-1}T^{-2})$ and $\mathrm{O}_p(N^{-3/2}T^{-3/2})$, respectively. Thus, the overall approximation error is $\mathrm{O}_p(N^{-1}T^{-3/2})$. A more accurate result, where, for instance, the approximation is correct up to $\mathrm{O}_p(N^{-1}T^{-1})$ terms can be obtained (at the expense of substantial analytic problems) by using a higher-order approximation for $(D)^{-1}$. This will not be pursued here. We use

$$\mathrm{E}(\hat{\delta} - \delta) = \mathrm{E}[(\bar{D})^{-1} - (\bar{D})^{-1}(D - \bar{D})(\bar{D})^{-1}]W'A\varepsilon + \mathrm{O}(N^{-1}T^{-3/2}), \tag{A.12}$$

and so our approximation to the bias equals

$$\begin{aligned}&\mathrm{E}[(\bar{D})^{-1} - (\bar{D})^{-1}(D - \bar{D})(\bar{D})^{-1}]W'A\varepsilon \\ &= 2(\bar{D})^{-1}\mathrm{E}W'A\varepsilon - (\bar{D})^{-1}\mathrm{E}[D(\bar{D})^{-1}W'A\varepsilon] \\ &= -2\sigma_\varepsilon^2(\iota_T' C \iota_T)\frac{N}{T}(\bar{D})^{-1}q - (\bar{D})^{-1}\mathrm{E}[D(\bar{D})^{-1}W'A\varepsilon], \end{aligned} \tag{A.13}$$

where we made use of (A.8). For the expression in square brackets in the second term we find that its expectation, i.e.,

$$\begin{aligned}\mathrm{E}[D(\bar{D})^{-1}W'A\varepsilon] = \mathrm{E}(&\bar{W}'A\bar{W} + \tilde{W}'A\bar{W} + \bar{W}'A\tilde{W} \\ &+ \tilde{W}'A\tilde{W})(\bar{D})^{-1}(\bar{W}'A\varepsilon + \tilde{W}'A\varepsilon),\end{aligned}$$

is the sum of the expectations of eight separate terms, viz.:

(i) $\quad \mathrm{E}\bar{W}'A\bar{W}(\bar{D})^{-1}\bar{W}'A\varepsilon = 0,$

(ii) $\quad \begin{aligned}\mathrm{E}\tilde{W}'A\bar{W}(\bar{D})^{-1}\bar{W}'A\varepsilon &= \mathrm{E}q\varepsilon'(I_N \otimes C'A_T)\bar{W}(\bar{D})^{-1}\bar{W}'A\varepsilon \\ &= \sigma_\varepsilon^2 \operatorname{tr}\{\bar{W}'(I_N \otimes A_T C A_T)\bar{W}(\bar{D})^{-1}\}q,\end{aligned}$

(iii) $$\mathrm{E}\bar{W}'A\tilde{W}(\bar{D})^{-1}\bar{W}'A\varepsilon = \mathrm{E}\bar{W}'(I_N \otimes A_T C)\varepsilon q'(\bar{D})^{-1}\bar{W}'A\varepsilon$$
$$= \mathrm{E}\bar{W}'(I_N \otimes A_T C)\varepsilon\varepsilon' A'\bar{W}(\bar{D})^{-1}q$$
$$= \sigma_\varepsilon^2 \bar{W}'(I_N \otimes A_T C A_T)\bar{W}(\bar{D})^{-1}q,$$

(iv) $$\mathrm{E}\tilde{W}'A\tilde{W}(\bar{D})^{-1}\bar{W}'A\varepsilon = 0 \qquad \text{(third moment)},$$

(v) $$\mathrm{E}\bar{W}'A\bar{W}(\bar{D})^{-1}\tilde{W}'A\varepsilon = -\sigma_\varepsilon^2(\iota_T' C \iota_T)\frac{N}{T}\bar{W}'A\bar{W}(\bar{D})^{-1}q,$$

(vi) $$\mathrm{E}\tilde{W}'A\bar{W}(\bar{D})^{-1}\tilde{W}'A\varepsilon = 0 \qquad \text{(third moment)},$$

(vii) $$\mathrm{E}\bar{W}'A\tilde{W}(\bar{D})^{-1}\tilde{W}'A\varepsilon = 0 \qquad \text{(third moment)},$$

(viii) $$\mathrm{E}\tilde{W}'A\tilde{W}(\bar{D})^{-1}\tilde{W}'A\varepsilon = \mathrm{E}q\varepsilon'(I_N \otimes C'A_T)(I_N \otimes A_T C)\varepsilon q'(\bar{D})^{-1}$$
$$\times q\varepsilon'(I_N \otimes C'A_T)\varepsilon$$
$$= \sigma_\varepsilon^4 N q'(\bar{D})^{-1} q\Bigg[-\frac{N}{T}(\iota_T' C\iota_T)\operatorname{tr}\{C'A_T C\}$$
$$+ 2\operatorname{tr}\{C'A_T C A_T C\}\Bigg]q.$$

Substitution of these terms in (A.13) and (A.12) respectively yields the result of the theorem. ■

Proof of Theorem 2. We have

$$\mathrm{E}\Lambda^{-1}(\hat{\delta} - \delta) = \mathrm{E}(\Lambda W'AW\Lambda)^{-1}\Lambda W'A\varepsilon$$
$$= \mathrm{E}[(\bar{B})^{-1} - (\bar{B})^{-1}(B - \bar{B})(\bar{B})^{-1}]\Lambda W'A\varepsilon + \mathrm{O}(N^{-1}T^{-3/2})$$
$$= -\sigma_\varepsilon^2 \Lambda^{-1}(\bar{D})^{-1}\Bigg(\frac{N}{T}(\iota_T' C\iota_T)[2q - \bar{W}'A\bar{W}(\bar{D})^{-1}q]$$
$$+ \operatorname{tr}\{\bar{W}'(I_N \otimes A_T C A_T)\bar{W}(\bar{D})^{-1}\}q + W'(I_N \otimes A_T C A_T)\bar{W}(\bar{D})^{-1}q$$
$$+ \sigma_\varepsilon^2 N q'(\bar{D})^{-1} q\Bigg[-\frac{N}{T}(\iota_T' C\iota_T)\operatorname{tr}\{C'A_T C\}$$
$$+ 2\operatorname{tr}\{C'A_T C A_T C\}\Bigg]q\Bigg) + \mathrm{O}(N^{-1}T^{-3/2}).$$

Premultiplication by Λ shows the [for $d(i) > 0$ increased] accuracy of the same approximation for $\mathrm{E}(\hat{\delta} - \delta)$ as given in Theorem 1. ■

Appendix B

Data according to (1) through (7), where the scalar x_{it} ($K = 1$) follows an AR(1) process, are simulated as follows. First we generate the latent variables v_{it} given in (10). Since data for all individuals are generated equivalently (but independently) we omit the index i, hence:

$$v_t = \gamma v_{t-1} + \beta x_t + \varepsilon_t, \qquad t = 1, \dots, T, \tag{B.1}$$

$$x_t = \rho x_{t-1} + \xi_t, \qquad t = 1, \dots, T, \tag{B.2}$$

where ε_t and ξ_t are mutually independent white noise series, with $\varepsilon_t \sim \mathrm{N}(0, \sigma_\varepsilon^2)$ and $\xi_t \sim \mathrm{N}(0, \sigma_\xi^2)$. Initial observations v_0 and x_0 are obtained from a procedure as in Kiviet (1986), which avoids 'warming-up' truncation errors.

Let $\mathscr{L}$ denote the lag operator then, upon substituting (B.2) in (B.1), we have $(1 - \gamma\mathscr{L})v_t = \beta(1 - \rho\mathscr{L})^{-1}\xi_t + \varepsilon_t$ or

$$\begin{aligned} v_t &= \beta[(1 - \gamma\mathscr{L})(1 - \rho\mathscr{L})]^{-1}\xi_t + (1 - \gamma\mathscr{L})^{-1}\varepsilon_t \\ &= \beta\varphi_t + \psi_t. \end{aligned} \tag{B.3}$$

Hence, v_t is the weighted sum of mutually independent stationary processes φ_t and ψ_t, which are an AR(2) and an AR(1) process, respectively.

Data for x_t and v_t (i.e., φ_t and ψ_t) can be obtained (from $t = 0$ onwards) from drawings $(\xi_0, \dots, \xi_T)$ and $(\varepsilon_0, \dots, \varepsilon_T)$ as follows:

$$x_0 = \xi_0(1 - \rho^2)^{-1/2}, \quad x_t = \rho x_{t-1} + \xi_t, \qquad t = 1, \dots, T, \tag{B.4}$$

$$\psi_0 = \varepsilon_0(1 - \gamma^2)^{-1/2}, \quad \psi_t = \gamma\psi_{t-1} + \varepsilon_t, \qquad t = 1, \dots, T. \tag{B.5}$$

In a similar way the AR(2) series is obtained from the scheme

$$\begin{aligned} \varphi_0 &= \xi_0[\mathrm{var}(\varphi^t)]^{1/2}, \\ \varphi_1 &= \varphi_0 \mathrm{cor}(\varphi_t, \varphi_{t-1}) + \xi_1[\mathrm{var}(\varphi_t)]^{1/2}\{1 - [\mathrm{cor}(\varphi_t, \varphi_{t-1})]^2\}^{1/2}, \\ \varphi_t &= (\gamma + \rho)\varphi_{t-1} - \gamma\rho\varphi_{t-2} + \xi_t, \qquad t = 2, \dots, T, \end{aligned} \tag{B.6}$$

where

$$\begin{aligned} \mathrm{cor}(\varphi_t, \varphi_{t-1}) &= \mathrm{cov}(\varphi_t, \varphi_{t-1})/\mathrm{var}(\varphi_t) = (\gamma + \rho)/(1 + \gamma\rho), \\ \mathrm{cor}(\varphi_t, \varphi_{t-2}) &= \mathrm{cov}(\varphi_t, \varphi_{t-2})/\mathrm{var}(\varphi_t) = (\gamma + \rho)^2/(1 + \gamma\rho) - \gamma\rho, \\ \mathrm{var}(\varphi_t) &= \sigma_\xi^2[1 - (\gamma + \rho)\mathrm{cor}(\varphi_t, \varphi_{t-1}) + \gamma\rho\,\mathrm{cor}(\varphi_t, \varphi_{t-2})]^{-1}. \end{aligned} \tag{B.7}$$

From (9) and (B.3) we find

$$y_t = \beta\varphi_t + \psi_t + \eta/(1 - \gamma), \tag{B.8}$$

and the dependent variable can be generated according to this scheme, upon using (B.5) through (B.7).

References

Ahn, S.C. and P. Schmidt, 1993, Efficient estimation of models for dynamic panel data, Mimeo.

Anderson, T.W. and C. Hsiao, 1981, Estimation of dynamic models with error components, Journal of the American Statistical Association 76, 589–606.

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

Arellano, M., 1989, A note on the Anderson–Hsiao estimator for panel data, Economics Letters 31, 337–341.

Arellano, M. and S. 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.

Baltagi, B.H. and B. Raj, 1992, A survey of recent theoretical developments in the econometrics of panel data, Empirical Econometrics 17, 85–109.

Beggs, J.J. and M. Nerlove, 1988, Biases in dynamic models with fixed effects, Economics Letters 26, 29–31.

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

Blundell, R. and R.J. Smith, 1991, Conditions initiales et estimation efficace dans les modèles dynamiques sur données panel: Une application au comportement d'investissement des entreprises, Annales d'Economie et de Statistique 20/21, 109–123.

Doel, I.T. van den and J.F. Kiviet, 1993, Neglected dynamics in panel data models: Consequences and detection in finite samples, Statistica Neerlandica, forthcoming.

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

Hsiao, C., 1985, Benefits and limitations of panel data, Econometric Reviews 4, 121–174 (with comments and reply, 175–189).

Hsiao, C., 1986, Analysis of panel data (Cambridge University Press, Cambridge).

Kiviet, J.F., 1986, On the rigour of some misspecification tests for modelling dynamic relationships, Review of Economic Studies LIII, 241–261.

Kiviet, J.F. and G.D.A. Phillips, 1993, Alternative bias approximations in regressions with a lagged dependent variable, Econometric Theory 9, 62–80.

Kiviet, J.F. and G.D.A. Phillips, 1994, Bias assessment and reduction in linear error-correction models, Journal of Econometrics 63, 215–243.

Maddala, G.S., 1971, The use of variance components models in pooling cross section and time series data, Econometrica 39, 341–358.

Nankervis, J.S. and N.E. Savin, 1987, Finite sample distributions of t and F statistics in an AR(1) model with an exogenous variable, Econometric Theory 3, 387–408.

Nerlove, M., 1971, Further evidence on the estimation of dynamic economic relations from a time series of cross-sections, Econometrica 39, 359–387.

Nerlove, M. and P. Balestra, 1992, Formulation and estimation of econometric models for the analysis of panel data, in: L. Màtyàs and P. Sevestre, eds., The econometrics of panel data (Kluwer Academic Publishers, Dordrecht).

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

Sevestre, P. and A. Trognon, 1983, Propriétés de grands echantillons d'une classe d'estimateurs des modèles autorégressifs à erreurs composées, Annales de l'INSEE 50, 25–48.

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

Trognon, A, 1978, Miscellaneous asymptotic properties of OLS and ML estimators in dynamic error components models, Annales de l'INSEE 30, 631–658.