Bun, Maurice J.G.; Kiviet, Jan F.

Working Paper

The Accuracy of Inference in Small Samples of Dynamic Panel Data Models

Tinbergen Institute Discussion Paper, No. 01-006/4

Provided in Cooperation with:
Tinbergen Institute, Amsterdam and Rotterdam

Suggested Citation: Bun, Maurice J.G.; Kiviet, Jan F. (2001) : The Accuracy of Inference in Small Samples of Dynamic Panel Data Models, Tinbergen Institute Discussion Paper, No. 01-006/4, Tinbergen Institute, Amsterdam and Rotterdam

This Version is available at:
https://hdl.handle.net/10419/86074

TI 2001-006/4
Tinbergen Institute Discussion Paper

The Accuracy of Inference in Small Samples of Dynamic Panel Data Models

Maurice J.G. Bun
Jan F. Kiviet

Tinbergen Institute
The Tinbergen Institute is the institute for economic research of the Erasmus Universiteit Rotterdam, Universiteit van Amsterdam and Vrije Universiteit Amsterdam.

Tinbergen Institute Amsterdam
Keizersgracht 482
1017 EG Amsterdam
The Netherlands
Tel.: +31.(0)20.5513500
Fax: +31.(0)20.5513555

Tinbergen Institute Rotterdam
Burg. Oudlaan 50
3062 PA Rotterdam
The Netherlands
Tel.: +31.(0)10.4088900
Fax: +31.(0)10.4089031

The Accuracy of Inference in Small Samples of Dynamic Panel Data Models: Simulation Evidence and Empirical Results*

Maurice J.G. Bun and Jan F. Kiviet
(University of Amsterdam)

9 January 2001

JEL-code: C13; C23; R30

Keywords: bias correction, dynamic panel data model, Monte Carlo simulation, small sample accuracy, test size, variance estimation

Abstract

Through Monte Carlo experiments the small sample behavior is examined of various inference techniques for dynamic panel data models when both the time-series and cross-section dimensions of the data set are small. The LSDV technique and corrected versions of it are compared with IV and GMM regarding: coefficient bias, accuracy of variance estimators - both of the disturbances and of the coefficient estimators - and the actual size of coefficient tests. A reasonably simple and consistent bias adjusted LSDV estimator, for which we find an analytical and a bootstrap consistent estimator of its variance, performs relatively well. Further higher-order refinements of the bias correction do not improve the accuracy considerably. Most techniques show substantial size distortions for asymptotic t tests. Finally, it is illustrated how these findings help to interpret empirical results on the relationship between so-called dynamic externalities and local economic activity in Moroccan urban areas.

1. Introduction

Least-squares estimation techniques for static panel data models give consistent estimators in a dynamic setting only when the number of time observations T in

* Tinbergen Institute & Faculty of Economics and Econometrics, University of Amsterdam, Roeterstraat 11, 1018 WB Amsterdam, The Netherlands (MBUN@FEE.UvA.NL and JFK@FEE.UvA.NL)

the data set approaches infinity. Since in micro-economics the typical dimension of a panel data set is a short time span for a large cross-section various authors have paid attention to this problem and have proposed IV and GMM alternatives, which are consistent for fixed T and an infinite number of cross-section observations N (Anderson and Hsiao, 1982; Arellano and Bond, 1991; Ahn and Schmidt, 1995). GMM estimators typically use more orthogonality conditions than their simple IV counterparts and they take the covariance structure of the disturbances into account. Therefore they are asymptotically more efficient. In an attempt to avoid the problem of weak instruments, Kiviet (1995) follows another route to achieve large N consistent estimation and to enhance efficiency. An asymptotic approximation formula for the finite sample bias of the inconsistent LSDV estimator is exploited to remove inconsistency and to reduce mean squared error. In Kiviet (1999) this approximation formula has been slightly refined by taking a particular further higher-order term into account, but its actual performance has not been examined yet.

Although IV and especially GMM estimators have attractive asymptotic properties, Monte Carlo experiments (Arellano and Bond, 1991; Kiviet, 1995; Harris and Mátyás, 1996; Judson and Owen, 1999) have shown that for these techniques the quality of the asymptotic approximations in finite samples depends heavily on the actual parameter values of the model and on the dimensions of the available data set. In most of these simulation experiments short time series and reasonably large cross-section samples have been examined and attention is focused on bias and mean squared error of the coefficient estimators. Not much is known yet (for an exception, see Judson and Owen, 1999) about the performance of the various techniques when both dimensions are small, e.g. $N = 10$, $T = 10$ or $N = 5$, $T = 20$. Such dimensions can be relevant in both macro- and micro-economic applications, where the time span is one or two decades and the cross-section consists of a small number of countries or sectors. Although some attention has been paid to assessing the accuracy of variance estimators (Arellano and Bond, 1991), little is known about the actual reliability of asymptotic test procedures in panels of a limited sample size.

The purpose of this paper is to produce further insights into the finite sample properties of various inference techniques for dynamic panel data models, especially when both dimensions of the data set are small. By simulation we examine for this situation the bias and mean squared error of various coefficient estimators, the bias in related estimators for the disturbance variance and the accuracy of estimators of the coefficient standard errors. Also we examine the actual size of simple asymptotic coefficient tests. More in particular, we will focus on various possible implementations of the bias corrected LSDV estimator (LSDVc) as proposed by Kiviet (1995, 1999). First, we will analyze the magnitude of the con-

tributions of the separate terms in the bias approximation. Second, as the bias approximation depends on the unknown model parameters, we also investigate the sensitivity of the LSDVc estimator with respect to the choice of the preliminary consistent estimator. Third, we develop an appropriate estimator for the standard error of the LSDVc estimator, which proves to be nontrivial for the small T, large N case. We examine whether it is possible to exploit such an estimator in tests on coefficient values. Here, we also analyze the accuracy of estimators and tests in dynamic panel data models including non-stationary variables, since this situation is likely to arise in applied research.

In a rather extensive Monte Carlo study Kiviet (1995) examines the bias and mean squared error of the corrected LSDV estimator and compares these with IV and GMM alternatives for the small T, relatively large N panel with stationary variables. One conclusion is that in samples of such a size there is no superior technique yet over a broad range of relevant parameter values of the model. More recent simulation results of Harris and Mátyás (1996) and Judson and Owen (1999) indicate, however, that especially in samples with smaller values for N the semi-consistent IV and GMM techniques often perform rather poorly as far as the bias and efficiency of coefficient estimators are concerned. Regarding variance estimators the simulation results of Arellano and Bond (1991) show that in the panel with small T and large N the asymptotic approximation of the standard errors of their two-step GMM estimator is biased downward, but that one-step estimators of the asymptotic standard errors are quite accurate. In this paper less favorable results for the one-step GMM estimator are obtained when both N and T are small. Often the simulation results show a substantial downward bias for estimated asymptotic standard errors of coefficient estimators. We also find that for all procedures presently available the true size of t-type tests may differ substantially from the nominal level.

The outline of this study is as follows. In Section 2 the estimators to be analyzed are introduced. A review of their asymptotic distributions and an expression for an estimator for the variance of corrected LSDV estimators, derived in an Appendix, is given in Section 3. In Section 4 the simulation design is described, with further details in another Appendix, and the finite sample behavior of the various estimators and tests are compared. Section 5 contains an application of the various procedures in a panel analysis of so-called dynamic externalities and local economic activity in Morocco, and Section 6 concludes.

2. Estimators for dynamic panel data models

We consider estimation methods for the standard linear first-order dynamic panel data model

$$y_{it} = \gamma y_{i,t-1} + \beta' x_{it} + \eta_i + \varepsilon_{it}. \qquad (i = 1, ..., N;\ t = 1, ..., T) \qquad (2.1)$$

In this model the dependent variable y_{it} is regressed on a vector of K exogenous explanatory variables x_{it} and the one period lagged value of the dependent variable. The composite disturbance term $u_{it} = \eta_i + \varepsilon_{it}$ consists of an individual specific effect η_i, which is either fixed or random, and a white noise term ε_{it} with variance σ_ε^2. In case of random effects, η_i has mean zero and finite variance σ_η^2, is uncorrelated between units i and not correlated with the general disturbance term ε_{it}. Stacking the observations over time and across individuals one gets

$$y = W\delta + (I_N \otimes \iota_T)\eta + \varepsilon, \qquad (2.2)$$

where $\delta = (\gamma, \beta')'$, y and $W = [y_{-1} \vdots X]$ are $NT \times 1$ and $NT \times (K+1)$ matrices of stacked observations, ε is the $NT \times 1$ vector of disturbances and $\iota_T = (1, ..., 1)'$ a $T \times 1$ vector of ones.

2.1. LSDV estimator

It is clear from (2.1) that when the effects are random, the regressor $y_{i,t-1}$ is correlated with the composite disturbance term $\eta_i + \varepsilon_{it}$. Hence, standard estimators for the random effects model are inconsistent (results on the asymptotic bias can be found in Sevestre and Trognon, 1985). Therefore, one usually avoids to estimate the coefficients $(\gamma, \beta')'$ by error component techniques, because this leads to many complications (see Sevestre and Trognon, 1996). It is more common to assume that the individual effects are fixed, or otherwise to treat genuinely random individual effects as fixed anyhow, in order to deal with their likely non-zero correlation with the regressors. Then the N components of the vector $\eta = (\eta_1, ..., \eta_N)'$ constitute N unknown parameters corresponding to the dummy variables $I_N \otimes \iota_T$ in (2.2). Note that their estimation will lead to a considerable loss of degrees of freedom, especially when N is large. Estimation of the $N + K + 1$ coefficients of (2.2) by ordinary least squares yields estimates which are called Least Squares Dummy Variables (LSDV) or fixed effect estimates. Using standard partitioned regression results, the resulting estimator for δ can be expressed as

$$\begin{aligned} \hat{\delta}_{LSDV} &= (W'AW)^{-1}W'Ay \qquad (2.3) \\ &= \delta + (W'AW)^{-1} \begin{pmatrix} y'_{-1}A\varepsilon \\ X'A\varepsilon \end{pmatrix}, \end{aligned}$$

where the $NT \times NT$ matrix $A = I_N \otimes (I_T - \frac{1}{T}\iota_T \iota_T')$. Note that A is the within transformation which wipes out the individual effects. For ease of exposition it is assumed that all the explanatory variables are time variant so that $W'AW$ is invertible. Although exogeneity of the regressors X implies $E(X'A\varepsilon) = 0$, we have $E(W'A\varepsilon) \neq 0$, since it can be shown that

$$E(y_{-1}'A\varepsilon) = -\sigma_\varepsilon^2 \frac{N}{1-\gamma}\left(1 - \frac{1-\gamma^T}{T(1-\gamma)}\right). \tag{2.4}$$

Therefore, the LSDV estimator of δ is consistent for $T \to \infty$ but inconsistent for $N \to \infty$ and T finite, see Nickell (1981).

2.2. IV estimation

A different transformation of (2.2) for removing the individual specific effects is first differencing, i.e.

$$Dy = \gamma Dy_{-1} + \beta' DX + D\varepsilon. \tag{2.5}$$

Here $D = I_N \otimes D_T$ is the $N(T-1) \times NT$ matrix operator with the $(T-1) \times T$ matrix D_T defined as

$$D_T = \begin{bmatrix} -1 & 1 & 0 & \cdot & \cdot & 0 \\ 0 & -1 & 1 & \cdot & & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & & \cdot & -1 & 1 & \cdot \\ 0 & \cdot & \cdot & 0 & -1 & 1 \end{bmatrix}.$$

Since $E(y_{-1}'D'D\varepsilon) = -E(y_{-1}'\varepsilon_{-1}) = -E(\varepsilon_{-1}'\varepsilon_{-1}) = -\sigma_\varepsilon^2 N(T-1)$ least-squares is inconsistent now, irrespective of how the sample size is extended. However, Anderson and Hsiao (1982) propose two simple instrumental variables (IV) estimators for the parameter vector δ, which make use of either Dy_{-2} or y_{-2} as instrument for Dy_{-1}. The proposed instruments are uncorrelated with the disturbance term $D\varepsilon$ and correlated with Dy_{-1}, so they are valid. In this study we will analyze the second option only, i.e. using y_{-2} as instrument. Then estimation of (2.5) leads to the estimator

$$\hat{\delta}_{IV} = (Z'W^*)^{-1}Z'Dy, \tag{2.6}$$

where Z and W^* are $N(T-1) \times (K+1)$ matrices and Dy is a $N(T-1) \times 1$ vector. More in particular $W^* = [Dy_{-1} : DX]$, with DX a $NT(-1) \times K$ matrix and $Z = [y_{-2} : DX]$. Although the Anderson and Hsiao IV estimators, unlike $\hat{\delta}_{LSDV}$, are consistent for $N \to \infty$ and finite T, they are inefficient because (2.6) does not exploit all valid instruments neither takes into account that the disturbance term in (2.5) is an MA(1) process rather than white noise.

2.3. GMM estimation

Enhancement of efficiency can be pursued by a more elaborate estimation method, namely Generalized Method of Moments (Arellano and Bond, 1991; Ahn and Schmidt, 1995), which copes with both the correlation structure of the disturbances and the exploitation of any further instruments. The GMM estimator used in this paper is that of Arellano and Bond (1991). In their simulation study and empirical application they do not use all available overidentifying restrictions concerning the exogenous variables. Their $(T-1) \times (\frac{1}{2}T(T-1)+1)$ matrix Z_i^* of instruments for individual i, also used in this paper, is

$$Z_i^* = \begin{bmatrix} y_{i,0} & 0 & 0 & 0 & 0 & 0 & \cdots & 0 & \cdots & 0 & \Delta x_{i,2} \\ 0 & y_{i,0} & y_{i,1} & 0 & 0 & 0 & & 0 & \cdots & 0 & \Delta x_{i,3} \\ 0 & 0 & 0 & y_{i,0} & y_{i,1} & y_{i,2} & & & & & \\ \vdots & & & & & & \ddots & & & & \\ 0 & 0 & 0 & 0 & 0 & 0 & \cdots & y_{i,0} & \cdots & y_{i,T-2} & \Delta x_{i,T} \end{bmatrix}. \quad (2.7)$$

The moment conditions can be written conveniently as $E(Z^{*'}D\varepsilon) = 0$, where $Z^* = (Z_1^{*'}, ..., Z_N^{*'})'$. The GMM estimator of δ is based on the sample moments $\frac{1}{N}\sum_{i=1}^{N} Z_i^{*'} D_T \varepsilon_i = 0$ and is obtained as

$$\hat{\delta}_{GMM} = \arg\min_{\delta} \varepsilon' D' Z^* G Z^{*'} D\varepsilon, \quad (2.8)$$

where G is a weighting matrix. The optimal choice for G in (2.8) is to make it proportional to the inverse of $\sigma_\varepsilon^2 V = E[Z^{*'}D\varepsilon\varepsilon' D' Z^*]$, which is the covariance matrix of the moment conditions. The estimator in (2.8) is equivalent with the GLS estimator of δ in

$$Z^{*'}Dy = Z^{*'}W^*\delta + Z^{*'}D\varepsilon, \quad (2.9)$$

where $\sigma_\varepsilon^2 V$ is the variance matrix of the transformed disturbance term $Z^{*'}D\varepsilon$. Then the GMM estimator equals

$$\hat{\delta}_{GMM} = (W^{*\prime} Z^* \hat{V}^{-1} Z^{*'} W^*)^{-1} W^{*\prime} Z^* \hat{V}^{-1} Z^{*'} Dy. \quad (2.10)$$

The estimator of V used in this paper is $\hat{V} = \frac{1}{N} Z^{*'} H Z^*$, which is equivalent with the GMM1 estimator of Arellano and Bond (1991). The $N(T-1) \times N(T-1)$ matrix $H = I_N \otimes H_{T-1}$. Multiplied by the disturbance variance σ_ε^2 this is the variance matrix of $D\varepsilon$, which is an MA(1) process under the assumption that ε is white noise. The matrix $H_{T-1} = D_T D_T'$ has twos on the main diagonal, minus ones on the first subdiagonals and zeros elsewhere.

2.4. Corrected LSDV estimation

Unlike the more traditional estimators for static models, the proposed IV and GMM estimation methods have attractive asymptotic properties in dynamic models. These will be discussed more thoroughly in the next section. The focus in this paper, however, is on finite sample properties, which can be analyzed either by analytical finite sample approximations or by Monte Carlo experiments. Kiviet (1995) derives an approximation formula for the bias of the LSDV estimator (2.3) in the normal stationary dynamic panel data model such that the approximation error has a magnitude which is of order $O(N^{-1})$ and of order $O(T^{-2})$ at the same time. In Kiviet (1999) this analysis is extended and an approximation to the bias is produced which is not only accurate to order $O(T^{-1})$ again, but now it is accurate to order $O(N^{-1})$ as well. These expressions for the approximate bias are rather complicated, but here we shall show that they can be simplified considerably. In the simulations to follow we shall examine closely the actual magnitude of the various contributions to the finite sample bias of terms of decreasing orders with respect of both T and N. This will help to select an implementation of a bias corrected LSDV estimator with attractive properties in finite samples.

In order to employ the usual O and O_p notation unambiguously for panel data, where the sample has two dimensions, we put this notation into a wider context defined as follows: Let ϕ and ψ be real numbers. If the elements $a_{k,l}$ (where $k = 1, ..., K$ and $l = 1, ..., L$) of a non-stochastic $K \times L$ matrix A constitute sequences $\{a_{k,l}(i,t)\}$ for $i = 1, ..., N$ and $t = 1, ..., T$, where we allow either $N \to \infty$ or $T \to \infty$ or both, then $A = O(N^{\phi})$ indicates that there exists a finite constant $\bar{a}_{\phi} > 0$ such that $\left|N^{-\phi} a_{k,l}(i,t)\right| < \bar{a}_{\phi}$, $\forall k, l, i, t$. Similarly $A = O(T^{\psi})$ indicates that there exists a finite constant $\bar{a}_{\psi} > 0$ such that $\left|T^{-\psi} a_{k,l}(i,t)\right| < \bar{a}_{\psi}$, $\forall k, l, i, t$. In addition, $A = O(N^{\phi}T^{\psi})$ indicates that there exists a finite constant $\bar{a}_{\phi,\psi} > 0$ such that $\left|N^{-\phi}T^{-\psi} a_{k,l}(i,t)\right| < \bar{a}_{\phi,\psi}$, $\forall k, l, i, t$.

When $A = O(N^{\phi})$ we say that the elements of A are at most of order N^{ϕ}, which is equivalent with $O(N^{\phi}T^{0})$, i.e. the quantities $N^{-\phi} a_{k,l}$ are finite, even when T approaches its limit. More generally, $A = O(N^{\phi}T^{\psi})$ indicates that all the elements of A are at most of order $N^{\phi}T^{\psi}$.

In addition we define that when the matrix A contains random elements, $A = O_p(N^{\phi})$ indicates that for all $\varepsilon > 0$ there exists a finite constant $\bar{a}_{\varepsilon} > 0$ and a positive integer N_{ε} such that $\text{Prob}(\left|N^{-\phi} a_{k,l}(i,t)\right| > \bar{a}_{\varepsilon}) < \varepsilon$, $\forall k, l, t$ and $i > N_{\varepsilon}$. Similarly, $A = O_p(T^{\psi})$ indicates that for all $\varepsilon > 0$ there exists a finite constant $\bar{a}_{\varepsilon} > 0$ and a positive integer T_{ε} such that $\text{Prob}(\left|T^{-\psi} a_{k,l}(i,t)\right| > \bar{a}_{\varepsilon}) < \varepsilon$, $\forall k, l, i$ and $t > T_{\varepsilon}$. And finally, $A = O_p(N^{\phi}T^{\psi})$ indicates that for all $\varepsilon > 0$ there exists a finite constant $\bar{a}_{\varepsilon} > 0$ and positive integers N_{ε} and T_{ε} such that $\text{Prob}(\left|N^{-\phi}T^{-\psi} a_{k,l}(i,t)\right| > \bar{a}_{\varepsilon}) < \varepsilon$, $\forall k, l$ and $i > N_{\varepsilon}, t > T_{\varepsilon}$.

When $A = O_p(N^{\phi})$ we say that the elements of A are bounded in probability

by order N^{ϕ}, which is equivalent with $O_p(N^{\phi}T^0)$, and implying that $N^{-\phi}a_{k,l}$ is of order unity in probability, irrespective of the value of T. More generally, $A = O_p(N^{\phi}T^{\psi})$ indicates that the elements of A converge at least at rate $N^{\phi}T^{\psi}$.

The above terminology on rates of convergence and orders of magnitude of double-indexed deterministic and stochastic sequences of vectors or matrices will be employed now for estimators in panel data models, especially for the LSDV estimator. It has been shown in Kiviet (1995, 1999) that

$$\begin{aligned} E(\hat{\delta}_{LSDV} - \delta) &= E\left((W'AW)^{-1}W'A\varepsilon\right) \qquad (2.11)\\ &= c_1(T^{-1}) + c_2(N^{-1}T^{-1}) + c_3(N^{-1}T^{-2}) + O(N^{-2}T^{-2}), \end{aligned}$$

where

$$\begin{aligned} c_1(T^{-1}) &= QE(W'A\varepsilon) = O(T^{-1})\\ c_2(N^{-1}T^{-1}) &= -QE(W'AW - Q^{-1})Q[W'A\varepsilon - E(W'A\varepsilon)] = O(N^{-1}T^{-1})\\ c_3(N^{-1}T^{-2}) &= -QE\left([(W'AW - Q^{-1})Q]^2\right)E(W'A\varepsilon) = O(N^{-1}T^{-2}), \end{aligned}$$

with

$$\begin{aligned} Q &= [E(W'AW)]^{-1} = [\bar{W}'A\bar{W} + \sigma_{\varepsilon}^2 tr(\Pi'\Pi)e_1e_1']^{-1}\\ \bar{W} &= E(W) = [E(y_{-1})\vdots X]\\ E(y_{i,0}) &= y_{i,0}, \quad E(y_{i,t}) = \gamma E(y_{i,t-1}) + \beta' x_{i,t} \quad (i = 1, ..., N; t = 1, ..., T)\\ e_1 &= (1, 0, ..., 0)' \quad \text{has } K+1 \text{ elements}\\ \Pi &= AL\Gamma,\ \Gamma = I_N \otimes \Gamma_T,\ \Gamma_T = (I_T - \gamma L_T)^{-1},\ L = I_N \otimes L_T, \end{aligned}$$

where the $T \times T$ matrix L_T has ones on its first lower subdiagonal and all other elements equal to zero. Using the shorthand notation

$$q_1 = Qe_1 \qquad \text{and} \qquad q_{11} = e_1'q_1,$$

and following the derivations as presented in the papers referred to above, one obtains

$$\begin{aligned} c_1(T^{-1}) &= \sigma_{\varepsilon}^2 tr(\Pi)q_1 \qquad (2.12)\\ c_2(N^{-1}T^{-1}) &= -QE(W'AWQW'A\varepsilon) + QE(W'A\varepsilon) \qquad (2.13)\\ &= -\sigma_{\varepsilon}^2[Q\bar{W}'\Pi A\bar{W} + tr(Q\bar{W}'\Pi A\bar{W})I_{K+1}\\ &\qquad +2\sigma_{\varepsilon}^2 q_{11}tr(\Pi'\Pi\Pi)I_{K+1}]q_1\\ c_3(N^{-1}T^{-2}) &= QE(W'AWQW'AW)QE(W'A\varepsilon) - QE(W'A\varepsilon) \qquad (2.14)\\ &= \sigma_{\varepsilon}^4 tr(\Pi)\{2q_{11}Q\bar{W}'\Pi\Pi'\bar{W}q_1\\ &\qquad +[(q_1'\bar{W}'\Pi\Pi'\bar{W}q_1)\\ &\qquad\quad +q_{11}tr(Q\bar{W}'\Pi\Pi'\bar{W}) + 2tr(\Pi'\Pi\Pi'\Pi)q_{11}^2]q_1\}. \end{aligned}$$

In Kiviet (1995) only the first two terms (2.12) and (2.13) have been taken into account, so the approximation error is still $O(N^{-1}T^{-2})$. The more general result with approximation error of order $O(N^{-2}T^{-2})$ can already be found in Kiviet (1999) but here the expressions are simpler due to removing factors $\bar{W}'A\bar{W}$ by exploiting $\bar{W}'A\bar{W} = Q^{-1} - \sigma_\varepsilon^2 tr(\Pi'\Pi)e_1e_1'$.

Note that all components given above implicitly contain contributions which are in fact of smaller order than the order indicated. For example, upon exploiting the simple structure of Γ_T, as was also done in (2.4), we obtain for (2.12)

$$c_1(T^{-1}) = \sigma_\varepsilon^2 tr(\Pi)q_1 = -\sigma_\varepsilon^2 \left[\frac{N}{1-\gamma} - \frac{N}{T\,(1-\gamma)^2} + \frac{N\gamma^T}{T\,(1-\gamma)^2}\right] q_1, \qquad (2.15)$$

and so, since $q_1 = O(N^{-1}T^{-1})$, $c_1(T^{-1})$ also contains contributions of order $O(T^{-2})$ and smaller. Note that such is fully in accordance with our definitions given above, i.e. $O(T^{-1})$ refers to "is at most of order T^{-1}" (but could even be smaller, or be a combination of terms of order T^{-1} and smaller terms). The approximation to the LSDV bias established by the first of the three components of (2.15), i.e.

$$E(\hat{\delta}_{LSDV} - \delta) = -\sigma_\varepsilon^2 q_1 \frac{N}{1-\gamma} + O(T^{-2}) + O(N^{-1}T^{-1}), \qquad (2.16)$$

is an extremely simple one, which may work well for moderately large N and T, because the "pure" $O(T^{-1})$ contribution to the bias has been separated here from $O(N^{-1}T^{-1})$ and "pure" $O(T^{-2})$ contributions, together constituting a hybrid remainder term.

In numerical calculations for relevant data generating processes we shall examine the bias approximation

$$B_{LSDV}^{(0)} = -\sigma_\varepsilon^2 q_1 \frac{N}{1-\gamma}, \qquad (2.17)$$

and make comparisons with the approximations produced by

$$B_{LSDV}^{(j)} = \sum_{h=1}^{j} c_h, \qquad \text{for } j = 1, 2, 3, \qquad (2.18)$$

where c_h $(h = 1, 2, 3)$ refers to the contributions to the bias given in (2.12), (2.13) and (2.14) respectively. By assessing the true bias from Monte Carlo simulations we shall examine the relative accuracy of the various approximations and examine the importance of the inclusion of higher-order terms.

It is known from simulation studies that, despite its serious bias, the LSDV estimator has a relatively small variance as compared with IV and GMM techniques. Therefore, on basis of a mean squared error criterion, a reasonably efficient procedure may result when a bias approximation is used to correct for the bias in the LSDV estimator. We shall examine the implementations

$$\hat{\delta}^{(j)}_{LSDVc,h} = \hat{\delta}_{LSDV} - \hat{B}^{(j)}_{LSDV,h}, \qquad j = 0, .., 3, \tag{2.19}$$

where any consistent preliminary estimators $\hat{\delta}_h$ and $\hat{\sigma}^2_{\varepsilon,h}$, obtained from estimation technique $h = IV, GMM$, can be used to estimate the correction term.

Numerous other estimators have been proposed for the dynamic panel data model, and therefore the techniques described in this section are not exhaustive. See, for example, Harris and Mátyás (1996) for a broader overview of the existing techniques for the dynamic panel data model. However, the techniques discussed in this section are considered to be either the most common or the most promising ones.

3. Asymptotic properties

In this section we establish limiting distributions for the various estimators, either for $N \to \infty$ and T finite, or for $T \to \infty$ and N finite, and occasionally for $N \to \infty$, $T \to \infty$. We will not consider the limiting behavior of the estimators when both N and T grow, but at different rates (see Alvarez and Arellano, 1998). Note that an appropriate asymptotic variance for LSDVc estimators (2.19) has not been derived yet for the case $N \to \infty$ and T finite. The last subsection is about the asymptotic distribution of simple t-ratios for the various estimators.

3.1. IV and GMM estimators

First, we focus on the method of moment estimators. Under standard regularity conditions, and upon defining

$$\Sigma_{ZW^*} = \operatorname*{plim}_{N\to\infty} \frac{1}{N} Z'W^* \tag{3.1}$$

$$\Sigma_{ZHZ} = \operatorname*{plim}_{N\to\infty} \frac{1}{N} Z'HZ \tag{3.2}$$

$$\Sigma_{Z^*W^*} = \operatorname*{plim}_{N\to\infty} \frac{1}{N} Z^{*\prime}W^* \tag{3.3}$$

$$\Sigma_{Z^*HZ^*} = \operatorname*{plim}_{N\to\infty} \frac{1}{N} Z^{*\prime}HZ^*, \tag{3.4}$$

the usual asymptotic reasoning yields the normal limiting distributions for the IV and GMM estimators for fixed T, viz.

$$\sqrt{N}\left(\hat{\delta}_{IV}-\delta\right)\underset{N\to\infty}{\overset{d}{\longrightarrow}}\mathcal{N}\left[0,\ \sigma_{\varepsilon}^{2}\Sigma_{ZW^{*}}^{-1}\Sigma_{ZHZ}(\Sigma_{ZW^{*}}')^{-1}\right] \tag{3.5}$$

and

$$\sqrt{N}\left(\hat{\delta}_{GMM}-\delta\right)\underset{N\to\infty}{\overset{d}{\longrightarrow}}\mathcal{N}\left[0,\ \sigma_{\varepsilon}^{2}(\Sigma_{Z^{*}W^{*}}'\Sigma_{Z^{*}HZ^{*}}^{-1}\Sigma_{Z^{*}W^{*}})^{-1}\right]. \tag{3.6}$$

>From the above, we obtain the asymptotic approximations

$$\hat{\delta}_{IV}\underset{N\to\infty}{\overset{a}{\sim}}\mathcal{N}\left[\delta,\ \sigma_{\varepsilon}^{2}(Z'W^{*})^{-1}Z'HZ(W^{*\prime}Z)^{-1}\right] \tag{3.7}$$

and

$$\hat{\delta}_{GMM}\underset{N\to\infty}{\overset{a}{\sim}}\mathcal{N}\left[\delta,\ \sigma_{\varepsilon}^{2}(W^{*\prime}Z^{*}\hat{V}^{-1}Z^{*\prime}W^{*})^{-1}\right]. \tag{3.8}$$

Consistent estimators for σ_{ε}^{2}, the variance of the disturbance term ε, are given by

$$\hat{\sigma}_{\varepsilon,h}^{2}=\frac{\hat{u}_{h}'A\hat{u}_{h}}{N(T-1)-(K+1)},\quad(h=IV,\ GMM) \tag{3.9}$$

where

$$\hat{u}_{h}=y-W\hat{\delta}_{h}. \tag{3.10}$$

Note that the numerator in (3.9) is the sum of squares of the group-demeaned (in order to account for the estimated individual effect coefficients) residuals, and the denominator contains a degrees of freedom correction for the $N+K+1$ estimated coefficients. Obviously, $\hat{\sigma}_{\varepsilon,IV}^{2}$ and $\hat{\sigma}_{\varepsilon,GMM}^{2}$ are semi-consistent, like $\hat{\delta}_{IV}$ and $\hat{\delta}_{GMM}$, for $N\to\infty$ and any T.

3.2. LSDV estimator

To obtain similar results for the LSDV estimator, where

$$\left(\hat{\delta}_{LSDV}-\delta\right)=\left(W'AW\right)^{-1}W'A\varepsilon, \tag{3.11}$$

we have to be a little bit more careful. We will obtain results for $T\to\infty$ and N finite and $N\to\infty$ and T finite. The right-hand expression of (3.11) has two factors. First, we will consider the order of the expectations of both $W'AW$ and $W'A\varepsilon$, which have already been presented below (2.11). They are obtained from

$$AW=E(AW)+[AW-E(AW)]=A\bar{W}+\Pi\varepsilon e_{1}', \tag{3.12}$$

where $\Pi=I_{N}\otimes\Pi_{T}$ with $\Pi_{T}=A_{T}L_{T}\Gamma_{T}$. This yields

$$E(W'AW)=Q^{-1}=\bar{W}'A\bar{W}+\sigma_{\varepsilon}^{2}tr(\Pi'\Pi)e_{1}e_{1}'=O(NT) \tag{3.13}$$

and

$$\begin{aligned} E(W'A\varepsilon) &= E(\varepsilon'\Pi\varepsilon)e_1 = \sigma_\varepsilon^2 tr(\Pi)e_1 \\ &= E[\varepsilon'(I_N \otimes \Pi_T)\varepsilon]e_1 = \sigma_\varepsilon^2 N tr(\Pi_T)e_1 = O(N), \end{aligned} \tag{3.14}$$

because

$$tr(\Pi_T) = -\left(\frac{1}{1-\gamma} - \frac{1-\gamma^T}{T(1-\gamma)^2}\right), \tag{3.15}$$

see (2.15). From (3.12) we also find

$$\begin{aligned} \mathrm{Var}(W'A\varepsilon) &= \mathrm{Var}(\bar{W}'A\varepsilon + \varepsilon'\Pi\varepsilon e_1) \\ &= \sigma_\varepsilon^2 \bar{W}'A\bar{W} + \sigma_\varepsilon^4 N\left[tr\left(\Pi_T'\Pi_T\right) + tr\left(\Pi_T\Pi_T\right)\right]e_1 e_1' = O(NT). \end{aligned} \tag{3.16}$$

Now we define both

$$\Sigma_{WAW} = \operatorname*{plim}_{N\to\infty} \frac{1}{N} W'AW \tag{3.17}$$

for T finite, and

$$\Upsilon_{WAW} = \operatorname*{plim}_{T\to\infty} \frac{1}{T} W'AW \tag{3.18}$$

for N finite. First consider

$$\operatorname*{plim}_{T\to\infty} (\hat{\delta}_{LSDV} - \delta) = \Upsilon_{WAW}^{-1} \operatorname*{plim}_{T\to\infty} \frac{1}{T} W'A\varepsilon.$$

Since it follows from (3.14) and (3.16) respectively that, for N finite, $E(\frac{1}{T}W'A\varepsilon) = O(T^{-1})$ and $\mathrm{Var}(\frac{1}{T}W'A\varepsilon) = O(T^{-1})$, we have $\operatorname*{plim}_{T\to\infty} \frac{1}{T}W'A\varepsilon = 0$. Hence,

$$\operatorname*{plim}_{T\to\infty} \hat{\delta}_{LSDV} = \delta, \tag{3.19}$$

i.e. $\hat{\delta}_{LSDV}$ is consistent for $T \to \infty$.

Employing an appropriate central limit theorem yields

$$\sqrt{T}\left(\hat{\delta}_{LSDV} - \delta\right) \xrightarrow[T\to\infty]{d} \mathcal{N}\left[0,\ \sigma_\varepsilon^2 \Upsilon_{WAW}^{-1}\right], \tag{3.20}$$

leading to the asymptotic approximate distribution

$$\hat{\delta}_{LSDV} \underset{T\to\infty}{\overset{a}{\sim}} \mathcal{N}\left[\delta,\ \sigma_\varepsilon^2 (W'AW)^{-1}\right]. \tag{3.21}$$

Also, for

$$\hat{\sigma}_{\varepsilon,LSDV}^2 = \frac{\hat{u}_{LSDV}' A \hat{u}_{LSDV}}{N(T-1) - (K+1)}, \tag{3.22}$$

with $\hat{u}_{LSDV} = y - W\hat{\delta}_{LSDV}$, we find

$$\operatorname*{plim}_{T\to\infty} \hat{\sigma}^2_{\varepsilon,LSDV} = \sigma^2_\varepsilon.$$

In principle, of course, these large T approximations will be useful only when T is large relative to N. It is straightforward to show that similar asymptotic results can be obtained for both $N \to \infty$ and $T \to \infty$.

However, for T finite and $N \to \infty$ the situation is much more complicated. It follows from (3.14) and (3.16) that $E(\frac{1}{N}W'A\varepsilon) = O(1)$ and $\mathrm{Var}(\frac{1}{N}W'A\varepsilon) = O(N^{-1})$. Hence, using $\operatorname*{plim}_{N\to\infty} \frac{1}{N}\left(W'A\varepsilon - E\left(W'A\varepsilon\right)\right) = 0$, we find

$$\begin{aligned}
\delta^* &= \operatorname*{plim}_{N\to\infty}\left(\hat{\delta}_{LSDV} - \delta\right) = \Sigma^{-1}_{WAW} \operatorname*{plim}_{N\to\infty} \frac{1}{N} W'A\varepsilon \qquad (3.23)\\
&= \Sigma^{-1}_{WAW} \operatorname*{plim}_{N\to\infty} \frac{1}{N}\left(W'A\varepsilon - E\left(W'A\varepsilon\right)\right) + \Sigma^{-1}_{WAW} \lim_{N\to\infty} \frac{1}{N} E(W'A\varepsilon)\\
&= -\sigma^2_\varepsilon \left(\frac{1}{1-\gamma} - \frac{1-\gamma^T}{T\left(1-\gamma\right)^2}\right) \Sigma^{-1}_{WAW} e_1 = O(1),
\end{aligned}$$

which defines δ^* as the inconsistency (asymptotic bias) of LSDV for $N \to \infty$ and finite T, earlier addressed in Nickell (1981) and Sevestre and Trognon (1996).

It is interesting to examine whether, in addition to (3.20), the LSDV estimator has a limiting distribution for $N \to \infty$ and fixed T, and what the situation is for LSDVc estimators like (2.19). From (3.16) and (3.13) we obtain

$$\mathrm{Var}(W'A\varepsilon) = \sigma^2_\varepsilon Q^{-1} + \sigma^4_\varepsilon N tr\left(\Pi_T \Pi_T\right) e_1 e_1', \qquad (3.24)$$

and with (3.17) this yields

$$\frac{1}{\sqrt{N}}\left(W'A\varepsilon - E\left(W'A\varepsilon\right)\right) \xrightarrow[N\to\infty]{d} \mathcal{N}\left[0,\ \sigma^2_\varepsilon \Sigma_{WAW} + \sigma^4_\varepsilon tr(\Pi_T\Pi_T) e_1 e_1'\right]. \qquad (3.25)$$

Hence,

$$\begin{aligned}
&\left(\frac{1}{N}W'AW\right)^{-1}\left(\frac{1}{\sqrt{N}}W'A\varepsilon - \frac{1}{\sqrt{N}}E\left(W'A\varepsilon\right)\right)\\
&\qquad \xrightarrow[N\to\infty]{d} \mathcal{N}\left[0,\ \sigma^2_\varepsilon \Sigma^{-1}_{WAW} + \sigma^4_\varepsilon tr(\Pi_T\Pi_T)\Sigma^{-1}_{WAW} e_1 e_1' \Sigma^{-1}_{WAW}\right],
\end{aligned}$$

and, using (3.11) and (3.23), we obtain from the above

$$\begin{aligned}
&\sqrt{N}\left(\hat{\delta}_{LSDV} - \delta^* - \delta\right) \qquad (3.26)\\
&\qquad \xrightarrow[N\to\infty]{d} \mathcal{N}\left[0,\ \sigma^2_\varepsilon \Sigma^{-1}_{WAW} + \sigma^4_\varepsilon tr(\Pi_T\Pi_T)\Sigma^{-1}_{WAW} e_1 e_1' \Sigma^{-1}_{WAW}\right].
\end{aligned}$$

So, we find that for $N \to \infty$ and fixed T the estimator $\hat{\delta}_{LSDV}$ has a limiting distribution, but it is not centered at δ. Note that this limiting distribution depends not only on σ_ε^2 and Σ_{WAW}, but via Π_T also on γ. If δ^* were known, the corrected LSDV estimator $\hat{\delta}_{LSDV} - \delta^*$ would provide a $\sqrt{N}$-consistent estimator of δ with limiting distribution (3.26). Hence, even if δ^* were known and could also be used to obtain a consistent estimator $\hat{\sigma}^2$ for σ_ε^2, then $\hat{\sigma}^2(W'AW)^{-1}$ would not provide an adequate expression for the asymptotic variance. Thus, for $N \to \infty$ we find at least two striking differences with the $T \to \infty$ results: bias correction is indispensable for achieving consistency and the asymptotic variance of a consistent estimator is non-standard.

3.3. Corrected LSDV estimator

We will now analyze the limiting behavior of corrected LSDV estimators as proposed in (2.19). From (2.11) it follows that as long as $T \to \infty$ the limiting behavior of corrected LSDV estimators is equivalent to that of the ordinary LSDV estimator. Note that in this case also the ordinary LSDV estimator can be used as a preliminary consistent estimator of δ in the bias approximation, i.e. $h = LSDV$, IV, GMM in (2.19) yields consistency.

For T finite and $N \to \infty$ we have for the ordinary LSDV estimator the limiting result in (3.26). In practice δ^* of (3.23) is unknown, but we can estimate it. Employing a $\sqrt{N}$-consistent estimator $\hat{\delta}_h$, for instance $\hat{\delta}_{IV}$ of (3.5) or $\hat{\delta}_{GMM}$ of (3.6) and corresponding $\hat{\sigma}^2_{\varepsilon,h}$, yields

$$\hat{\delta}_h^* = -\hat{\sigma}^2_{\varepsilon,h} \left(\frac{1}{1-\hat{\gamma}_h} - \frac{1-\hat{\gamma}_h^T}{T\,(1-\hat{\gamma}_h)^2} \right) \left(\frac{1}{N} W'AW \right)^{-1} e_1, \quad h = IV,\ GMM. \tag{3.27}$$

We shall examine now the asymptotic distributions of the LSDVc estimators

$$\hat{\delta}_{LSDVc,h} = \hat{\delta}_{LSDV} - \hat{\delta}_h^*, \quad h = IV,\ GMM. \tag{3.28}$$

Note that this implementation corresponds to $\hat{\delta}^{(1)}_{LSDVc,h}$ given in (2.19). The higher order implementations with $j > 1$ have the same limiting distribution. To derive this limiting distribution we use

$$\sqrt{N}\left(\hat{\delta}_{LSDVc,h} - \delta\right) = \sqrt{N}\left(\hat{\delta}_{LSDV} - \delta^* - \delta\right) - \sqrt{N}\left(\hat{\delta}_h^* - \delta^*\right). \tag{3.29}$$

The first right-hand term has already been dealt with in (3.26) and in Appendix A the limiting distribution of the second term is derived and also of their difference. Since the second term has a finite limiting distribution indeed, i.e. is $O_p(1)$, the asymptotic variance of the corrected estimator with estimated δ^* is found to be

different from that where δ^* is known, which was already found to be non-standard itself.

In Appendix A we find

$$\sqrt{N}\left(\hat{\delta}_{LSDVc,h}-\delta\right)\xrightarrow[N\to\infty]{d}\mathcal{N}\left[0,\ \Omega_{LSDVc,h}\right],\qquad h=IV,\ GMM,$$

which leads to

$$\hat{\delta}_{LSDVc,h}\underset{N\to\infty}{\overset{a}{\sim}}\mathcal{N}\left[\delta,\ V_{LSDVc,h}\right],\qquad h=IV,\ GMM. \tag{3.30}$$

For $h=IV$ the expression for the asymptotic variance matrix has been derived and we find

$$\begin{aligned}V_{LSDVc,IV} &= \sigma_\varepsilon^2(W'AW)^{-1} \\ &\quad +\sigma_\varepsilon^4 N tr(\Pi_T\Pi_T)(W'AW)^{-1}e_1e_1'(W'AW)^{-1} \\ &\quad +\sigma_\varepsilon^4 N(W'AW)^{-1}W'AD'Z(W^{*\prime}Z)^{-1}e_1d_1' \\ &\quad +\sigma_\varepsilon^4 N d_1e_1'(Z'W^*)^{-1}Z'D'AW(W'AW)^{-1} \\ &\quad +\sigma_\varepsilon^4\left(6N/\left(T-1\right)\right)d_2d_2' \\ &\quad +\sigma_\varepsilon^6 N^2e_1'(Z'W^*)^{-1}Z'HZ(W^{*\prime}Z)^{-1}e_1d_1d_1' \\ &\quad +\sigma_\varepsilon^6 N^2 tr(\Pi_T\Gamma_T'L_T'K_T'D_T)(W'AW)^{-1}(W^{*\prime}Z)^{-1}e_1d_1' \\ &\quad +\sigma_\varepsilon^6 N^2 tr(\Pi_T\Gamma_T'L_T'K_T'D_T)d_1e_1'(Z'W^*)^{-1}(W'AW)^{-1},\end{aligned} \tag{3.31}$$

where

$$\begin{aligned}d_1 &= \left[\frac{1+\gamma^{T-1}}{\left(1-\gamma\right)^2}-\frac{2}{T}\frac{1-\gamma^T}{\left(1-\gamma\right)^3}\right](W'AW)^{-1}e_1 \\ d_2 &= tr(\Pi_T)(W'AW)^{-1}e_1.\end{aligned}$$

The expression in (3.31) can be used to estimate the asymptotic variance matrix of (3.30). For $h=GMM$ we did not derive such an expression due to the complexity of the dependence of the matrix of instruments Z^* on ε.

An alternative way to obtain empirical estimates of $V_{LSDVc,h}$, which avoids the use of complicated analytical expressions as in (3.31), is to bootstrap the standard errors. Exploiting the normality assumption, we may use a parametric bootstrap and the procedure for resampling is then as follows:

- Obtain the corrected LSDV estimators $\hat{\delta}_{LSDVc,h}$, $\hat{\sigma}^2_{\varepsilon,h}$ and $\hat{\eta}_{LSDVc,h}$;
- For $b=1,...,B$ generate a random sample $\varepsilon^{(b)}\sim\mathcal{N}[0,\ \hat{\sigma}^2_{\varepsilon,h}I_{NT}]$;

- Calculate $y^{(b)} = \hat{\gamma}_{LSDVc,h} y_{-1}^{(b)} + X\hat{\beta}_{LSDVc,h} + (I_N \otimes \iota_T)\hat{\eta}_{LSDVc,h} + \varepsilon^{(b)}$ where $y_{-1}^{(b)}$ contains the original starting values;
- Estimate the model with the resampled data to obtain the bootstrap LSDVc estimator $\hat{\delta}_{LSDVc,h}^{(b)}$.

Note that the presence of y_{-1} necessitates to use a recursive resampling scheme. The limiting distribution for the bootstrap estimator, conditional on $\hat{\delta}_{LSDVc,h}$, is

$$\sqrt{N}\left(\hat{\delta}_{LSDVc,h}^{(b)} - \hat{\delta}_{LSDVc,h}\right) \xrightarrow[N\to\infty]{d} \mathcal{N}\left[0,\ \Omega_{LSDVc,h}\right], \qquad h = IV,\ GMM.$$

The variance of $\hat{\delta}_{LSDVc,h}^{(b)}$ can be estimated from its empirical distribution by

$$\hat{V}_{LSDVc,h} = \frac{1}{B-1}\sum_{b=1}^{B}(\hat{\delta}_{LSDVc,h}^{(b)} - \bar{\delta}_{LSDVc,h})(\hat{\delta}_{LSDVc,h}^{(b)} - \bar{\delta}_{LSDVc,h})', \tag{3.32}$$

with $\bar{\delta}_{LSDVc,h} = \frac{1}{B}\sum_{b=1}^{B}\hat{\delta}_{LSDVc,h}^{(b)}$. Especially when N is large this variance estimator should be appropriate and may replace the estimation by the analytical expression (3.31).

3.4. t-ratios

The asymptotic distributions presented above can be used to construct asymptotic t-ratio tests for the null hypotheses $H_0 : \delta_k = \delta_k^0$ for $k = 1, ..., K+1$. Using e_k to denote a $K+1$ unit vector with all elements equal to zero, apart from a unit element at position k, the statistics can be expressed as ($j = 0, 1, 2, 3; \quad h = IV, GMM$):

$$t_{IV}(\delta_k^0) = \frac{e_k'\hat{\delta}_{IV} - \delta_k^0}{\sqrt{\hat{\sigma}_{\varepsilon,IV}^2 e_k' \left(Z'W^*\right)^{-1} Z'HZ \left(W^{*\prime}Z\right)^{-1} e_k}}, \tag{3.33}$$

$$t_{GMM}(\delta_k^0) = \frac{e_k'\hat{\delta}_{GMM} - \delta_k^0}{\sqrt{\hat{\sigma}_{\varepsilon,GMM}^2 e_k' W'Z^* \left(Z^{*\prime}HZ^*\right)^{-1} Z^{*\prime}We_k}}, \tag{3.34}$$

$$t_{LSDVc,h}^{(j)}(\delta_k^0) = \frac{e_k'\hat{\delta}_{LSDVc,h}^{(j)} - \delta_k^0}{\sqrt{e_k'\hat{V}_{LSDVc,h}e_k}}, \tag{3.35}$$

$$T_{LSDVc,h}^{(j)}(\delta_k^0) = \frac{e_k'\hat{\delta}_{LSDVc,h}^{(j)} - \delta_k^0}{\sqrt{\hat{\sigma}_{\varepsilon,LSDV}^2 e_k' \left(W'AW\right)^{-1} e_k}}, \tag{3.36}$$

$$T_{LSDV}(\delta_k^0) \;=\; \frac{e_k'\hat{\delta}_{LSDV} - \delta_k^0}{\sqrt{\hat{\sigma}^2_{\varepsilon,LSDV} e_k' \left(W'AW\right)^{-1} e_k}}. \tag{3.37}$$

Because they have been appropriately normalized, the statistics $t_{IV}(\delta_k^0)$, $t_{GMM}(\delta_k^0)$ and $t^{(j)}_{LSDVc,h}(\delta_k^0)$ will all converge to a standard normal distribution for $N \to \infty$ under the null hypothesis. For the tests $T^{(j)}_{LSDVc,h}(\delta_k^0)$, which combines the LSDVc coefficient estimator with the ordinary LSDV variance, and $T_{LSDV}(\delta_k^0)$ this is only the case when T approaches infinity.

4. Simulation results

The simulation design is basically the same as in Kiviet (1995), although slightly different values for some of the parameters have been chosen including cases with non-stationary regressors. Data for y have been generated according to equation (2.1) with $K = 1$. The generating equation for the explanatory variable x is

$$x_{it} = \rho x_{i,t-1} + \xi_{it} \qquad\qquad i = 1, ..., N;\ t = 1, ..., T, \tag{4.1}$$

where $\xi_{it} \sim \mathcal{IIN}[0, \sigma^2_\xi]$. As already noted in the introduction the focus is on samples in which both dimensions are relatively small. Therefore, the following three cases have been analyzed: $T = N = 10$; $T = 10$, $N = 20$ and $T = 20$, $N = 10$. Three values for γ (viz. 0.8; 0.5; 0.2) and two for ρ (viz. 0.8; 1.0) are considered. Each experiment consists of one thousand replications. The long-run effect $\beta/(1-\gamma)$ of x on y has been set equal to unity in all experiments. This implies that the impact multiplier β varies with the chosen values for γ. The variance of the disturbance term σ^2_ε is set at the value of one. So, for the three chosen sample sizes we consider six different designs concerning $\delta = (\gamma, \beta)'$, ρ and σ_ε.

Appendix B gives a detailed description on how the remaining design parameters σ_η and σ_ξ have been determined. By varying σ_η we control the relative impact on y of the two error components η and ε. The parameter σ_ξ has been determined by controlling the signal-to-noise ratio σ^2_s of the model. In Kiviet (1995) it has been shown that a proper comparison of simulation results over different parameter values requires to exercise control over some of the basic model characteristics such as the signal-to-noise ratio. Appendix B discusses also the modifications of the design required when the exogenous regressor is non-stationary. In this case ρ is equal to one, which is a situation not examined in Kiviet (1995). The signal-to-noise ratio $\sigma^2_{s,t}$ is now varying with t. Hence, we have chosen to fix the signal in the experiments at the mean value of $\sigma^2_{s,t}$ over the sample, and then the parameters σ_η and σ_ξ are determined as in the experiments with a stationary regressor x.

The following estimators have been analyzed: least squares dummy variables (LSDV), the Anderson and Hsiao (1982) simple IV estimator with the one-period lagged level as instrument (AHL), the GMM1 estimator of Arellano and Bond (1991) and the correction method of Kiviet (1995, 1999) based on either an initial GMM estimate (LSDVc) or IV estimate (LSDVc^{AHL}). In the correction formula we have always included the approximation $B^{(3)}_{LSDV}$, unless stated otherwise. Note that the uncorrected LSDV estimator is the only one that relies on asymptotics requiring large T.

In the experiments the GMM estimator is calculated for three sets of instruments. One, five and eight lagged values have been taken, which results in three different GMM estimators (GMMa, GMMb and GMMc respectively). The GMMc estimator is used as the preliminary estimator in the operational bias corrected estimator LSDVc.

We performed 1000 Monte Carlo replications. Results on the bias, variance and mean squared error of the various coefficient estimators are presented in Tables 1 and 2. To get a quick impression of the general qualities of the estimators (and to save space) we give here average outcomes over the three different γ values only (for particular T and N) but do report separate results for the stationary and non-stationary variable cases. We observe the following patterns in the simulation results. First, the sign and magnitude of the bias in estimating γ and β is for LSDV and the GMM procedures more or less comparable when both T and N are small. Second, the AHL estimator has almost negligible bias in estimating γ and a rather small bias for β, but it yields for both the largest variance. Regarding γ the large dispersion of the AHL estimator is confirmed by the results in the second column of Tables 1 and 2, which show the high frequency of estimates of γ outside the unit circle. Third, the LSDV procedure has relatively small variance as compared with GMM and IV. Fourth, bias correction reduces the bias considerably. Fifth, based on a mean squared error criterion the LSDVc procedures are never beaten by the other estimators. In addition, we note that the sensitivity with respect to the preliminary estimator used in the bias approximation, i.e. GMMc or AHL, is only marginal. Also when the difference in accuracy of AHL and GMM is large, the extent to which this is carried over to the accuracy of the bias corrected LSDV is minor. Finally, we note for GMM that its variance (but not its bias) decreases with the number of moment conditions employed.

In Tables 3 and 4 we compare the contributions of the various terms in the bias approximation for LSDV, calculated at the true parameter values, with the actual bias as estimated from the Monte Carlo for the stationary and non-stationary data respectively. We find that the magnitude of the bias depends much more on T than on N, and relatively little on ρ, the pattern of the regressor variable. The negative bias in the estimator of γ is larger in absolute value when γ is larger, and on the

whole it is much more serious than the bias in estimating β. In relative terms the bias in estimating γ is certainly serious when γ is small. The decomposition of the bias approximation in terms of different orders in T and N shows that the pure $O(T^{-1})$ approximation can be reasonably accurate when the bias is moderate but otherwise it may yield a considerable overstatement of the actual bias. Leaving into the approximation the contributions which are $o(T^{-1})$, but which are still $O(N^0)$ at the same time, is often profitable, but not always so. The contributions to the bias which are $O(N^{-1}T^{-1})$ and $O(N^{-1}T^{-2})$ are found to be of very limited actual magnitude, so one could decide to omit them from implementations of the LSDVc procedure.

We return now to results for the various estimation procedures and focus on variance estimates. In Table 5 results are presented for the case $\gamma = 0.5$ and $\rho = 0.8$. For reference, the first two columns give the bias of coefficient estimators in estimating γ and β for this specific experiment. Naturally, some of the results concerning the coefficient estimators will carry over to the findings for the disturbance variance estimators (3.9). The Table presents the estimated bias in the estimators of σ_ε as percentage of the true disturbance standard error, which is one in all experiments. Regarding this bias we find first of all that the magnitude of the bias of the LSDV and GMM estimators is more or less equal and always negative. Second, the bias in the AHL estimator is often positive. Third, bias problems seem to be less severe for larger T. In this Table the bias in estimating the standard deviation of the various coefficient estimators is also presented as percentage of the true standard deviation as estimated from the Monte Carlo. In many cases these relative biases of the coefficient standard deviation are worse than the bias of the disturbance variance (which is one of their determinants, of course). For LSDVc and LSDVc^{AHL} we present different variance estimators. The conventional large T expression of (3.21) is always reported. In addition, when $T \leqslant N$ also the "analytic" (only in case of LSDVc^{AHL}) and the bootstrap variance estimators as in (3.31) and (3.32) are presented. Especially for γ these alternative estimators seem more accurate than the standard expression. Similar results are obtained when allowing for a nonstationary regressor, i.e. $\rho = 1.0$. However, results not reported here show that for larger γ, i.e. $\gamma = 0.8$, the "analytic" variance estimator breaks down, while the bootstrap estimator is still relatively accurate.

We shall examine now what the implications are of the results on coefficient bias and bias in standard errors for the accuracy in small samples of the usual statistical inference as produced by t statistics. Table 6 shows that relying on asymptotic theory in small samples of the type analyzed here can be quite misleading as the empirical distribution of the t statistics can differ substantially from the asymptotic standard Normal. The true size for testing β is in many

cases still reasonably close to the nominal size. For testing γ all the techniques show substantial distortions. For AHL underrejection is more of a problem than overrejection. LSDV (understandably) but GMM too show huge distortions. The LSDVc results are generally much less extreme, but not really satisfactory. The differences in performance of LSDVc implementations when using alternative variance estimators seem to be marginal.

5. Dynamic externalities and local economic activity in Morocco

5.1. Introduction

In this section the performance of the various estimators will be examined in an empirical application. We analyse the determinants driving local economic activity in six major Moroccan urban areas. The existence of cities or urban areas is often explained by localization and urbanization economies, which arise as a result of a higher degree of both concentration and differentiation of economic activities. Localization economies emerge when similar firms cluster, while in case of urbanization economies it is the diversity of the industrial structure that matters. For example, localization economies permit firms to have better access to natural resources and lower transport costs, while urbanization economies enhance diversity of products and firms and increase market size.

While localization and urbanization economies describe the existing locational structure of a region, they do not necessarily explain the pattern of economic development through time. Growth theory (Romer, 1986) emphasizes the role of knowledge spillovers as an important source for technological change and hence economic growth. As close proximity of firms facilitates the transmission of ideas and innovations between firms, knowledge spillovers are most likely to occur in urban areas. The agglomeration economies arising from knowledge spillovers are called dynamic externalities. In contrast with traditional localization and urbanization economies, dynamic externalities explain both the formation of urban areas and local economic growth.

In the literature on dynamic externalities three main theories are distinguished. All these theories agree that knowledge spillovers are important, but they differ regarding their origins. First, Marshall-Arrow-Romer (MAR) externalities arise from intra industry knowledge spillovers, see for example Glaeser et al. (1992). MAR externalities imply that an increase in the concentration of firms of the same industry within a region facilitates knowledge spillovers, which in turn increases productivity. In other words, specialization of firms within a region will have positive effects on local economic activity. Another feature of MAR ex-

ternalities is that local monopoly rather than competition is better for growth. Second, contrary to MAR externalities, so called Jacobs (1969) externalities arise from inter industry knowledge spillovers or, in other words, diversity among firms is beneficial. Third, Porter (1990) externalities agree with MAR theory that a higher concentration of similar firms in a region facilitates knowledge spillovers. In contrast to MAR, however, Porter argues that a higher degree of local competition induces firms to innovate in order to remain competitive. In the view of Porter, which is supported also by Jacobs, competition is good for economic growth contrary to the prediction of MAR.

In Bun and El Makhloufi (1999) an attempt has been made to distinguish which type of externality, if any, is predominant for economic activity in some major urban areas in Morocco. As the empirical literature on dynamic externalities has focused exclusively on industrialized countries like the United States or European countries, this is a first attempt to quantify these theories for a developing country. It establishes significant MAR externalities, but provides mixed evidence on Jacobs and Porter effects.

An important question not addressed by Bun and El Makhloufi is to what extent the influence of dynamic externalities on local economic activity differs between manufacturing sectors. Glaeser et al. (1992) report estimation results on several subsamples of industries to check the robustness of their overall results. Henderson et al. (1995) argue that for large mature industrial sectors MAR externalities should be dominant, while for new industries also Jacobs externalities matter. Moreover, the degree of persistence in economic activity should be higher for large traditional industries as compared with smaller newer industries. Here, we will examine these features by assessing the degree of parameter heterogeneity between industries, i.e. separate regressions will be estimated for five main one-digit sectors available in the data. As these subsamples naturally have few observations in both the time and cross-section dimensions, the simulations in the previous section will be informative about the anomalous differences between estimates produced by the various techniques.

5.2. Data

The available data contain annual time series on several characteristics including gross value added, employment, wages and number of establishments. The period covered is 1985-1995 and the data are collected for eighteen two-digit manufacturing sectors and six large urban areas in Morocco. The data are obtained from the "Ministère de l'Industrie, du Commerce et de l'Artisanat" of Morocco.

In order to establish in an econometric analysis the importance of dynamic externalities for local economic activity the various types of externalities have

to be quantified. The data include three indicators, which aim to measure these externalities. Data about sectors $s = 1, ..., S$ and regions $r = 1, ..., R$ are available, so each cross-section unit i has a unique combination of s and r. We can identify each cross-section unit $i = 1, ..., SR$ uniquely with $i = (r-1)S + s$. Denote with emp_{srt} employment of industry s in region r at time period t. Furthermore, define emp_{st}, emp_{rt} and emp_t as employment at time t in sector s, region r or the whole country respectively.

Using the notation introduced above the definition of concentration or *specialization* becomes

$$sp_{it} = \frac{emp_{srt}/emp_{rt}}{emp_{st}/emp_t}. \tag{5.1}$$

This ratio measures the fraction of employment in sector s located in region r relative to the fraction of total employment in sector s of total employment in the country. Therefore, high levels for sp_{it} indicate that production of sector s is relatively concentrated in region r.

The measure for *diversity* is

$$dv_{it} = \sum_{k=1,k\neq s}^{S} \left[\frac{emp_{krt}}{emp_{rt} - emp_{krt}}\right]^2. \tag{5.2}$$

The ratio in this indicator is employment in sector k in region r relative to the total other manufacturing employment in region r. If this ratio is low for the majority of the sectors then there are many diversified activities in the region. Hence, a low level of dv_{it} implies a high degree of diversity.

The *competition* indicator is defined as

$$cp_{it} = \frac{ne_{srt}/emp_{srt}}{ne_{st}/emp_{st}}, \tag{5.3}$$

where ne_{srt} are the number of establishments of industry s in region r. If the number of establishments per worker for industry s and region r is relatively high to that of industry s in the whole country, then firms of sector s in that particular region are assumed to be relatively competitive.

The specialization and competition indicators given above are similar to those used in Glaeser et al. (1992). The diversity indicator is similar to the so-called Hirschman-Herfindahl index, which has been used also in Henderson et al. (1995) and other studies. Other explanatory variables measuring local market conditions are total regional manufacturing production (trp), which reflects market size[1], and

[1] To check the robustness of our results we have experimented with other measures, e.g. total regional population and total regional manufacturing employment. The estimation results of these specifications are similar to those presented here.

real unit wage costs ($wcap$). Finally, following De Lucio et al. (1998) real value added (va) will be used as a proxy for economic activity.

5.3. Estimation results

Because of the time series aspect of the data, a general dynamic specification has been estimated including as many lagged explanatory variables as seem required, including the one-period lagged value of the dependent variable. To account for unobserved heterogeneity, e.g. local resources or the institutional, cultural and political environment, individual specific effects have been included. The estimated specification is

$$\ln va_{it} = \gamma \ln va_{i,t-1} + \beta' x_{it} + \eta_i + \varepsilon_{it}, \tag{5.4}$$

with

$$x_{it} = (\ln trp_{it}, \ln trp_{i,t-1}, \ln wcap_{it}, \ln wcap_{i,t-1}, sp_{it}, sp_{i,t-1}, dv_{it}, dv_{i,t-1}, cp_{it}, cp_{i,t-1})'.$$

Complete data are available over the years 1985-1995 for 95 cross-section units. Bun and El Makhloufi (1999) report estimation results of (5.4) for the full sample. In that study, a significant positive effect for the specialization indicator has been found, while the coefficients of the diversity and competition indicators are negative. Because both Jacobs and Porter externalities should imply positive effects of competition, the results are in favour of MAR effects. An interesting question is whether these results hold for all sectors analysed. In other words, do the pooled estimates for the full sample blur sector specific differences?

We address this issue by dividing the full sample into five subsamples corresponding to the five one-digit manufacturing sectors and estimating specification (5.4) sector by sector. These sectors are food and agricultural products, textile and leather, chemical and allied products, metal and machinery, and electric and electronic manufacturing. The first three sectors are considerably larger than the latter two sectors. The cross-section dimensions of the subsamples are $N = \{18, 18, 29, 20, 10\}$ [2], while $T = 11$ in all cases. Two observations are lost in constructing the IV and GMM estimators and the number of regressors in (5.4) is $K = 11$.

Results not reported here suggest that the lag structure of specification (5.4) is adequate. Longer lags of both the dependent and explanatory variables were also tried, but in general these regressors do not seem to have a significant contribution in explaining $\ln(va)$. Regarding the estimators employed, apart from LSDV,

[2] For example, regarding the one-digit sector food and agricultural products we have $N = 18$ because data are available for three two-digit subsectors and six regions.

AHL and GMM results we present two different corrected LSDV estimators, i.e. based on preliminary GMM estimates (LSDVc) and AHL estimates (LSDVc^{AHL}) respectively. The corrected LSDV estimators use the approximation $B^{(3)}_{LSDV}$, i.e. $O(T^{-1})$, $O(N^{-1}T^{-1})$ and $O(N^{-1}T^{-2})$ bias terms have been removed. For LSDV, IV and GMM the estimation of standard errors is based on conventional asymptotic approximations, while for the LSDVc variants bootstrap standard errors will be reported.

To save space we present some selected estimation results only. Table 7 presents the estimation results for the sector chemical and allied products. Several remarks with respect to the pattern of the estimation results can be made which apply to the estimates of the other sectors too. First, regarding the autoregressive parameter all estimators produce more or less plausible estimates implying a stable autoregressive process for the dependent variable. Second, the GMM estimates of the autoregressive parameter are close to the LSDV estimate, while the bias corrected LSDV variants and especially the AHL estimator produce a relatively high estimate of the coefficient of the lagged dependent variable regressor. Third, the estimated standard errors of LSDV and LSDVc are in general considerably smaller than those of the AHL and GMM estimators. Especially the AHL estimates are often poorly determined. Fourth, while there are certainly differences between estimators, the pattern of the remaining estimates is often similar across estimators. An exception is the immediate effect of the diversity indicator, which shows differences between (corrected) LSDV and IV/GMM variants.

Regarding the estimation results for the three indicators of externalities (*sp*, *dv* and *cp*), the sign and significance of the corresponding coefficients are similar for all estimators. The impact multiplier or immediate effect of the specialization indicator is significant and positive, while the coefficient for the one-period lagged variable is often significant and negative. Regarding diversity the immediate effect is positive and its lagged effect significant and negative. Considering the competition indicator the immediate effect is significant and negative, while the lagged effect is positive in general.

Table 8 reports estimation results for specification (5.4) for each one-digit sector. Only the LSDVc estimates are shown, but the pattern for the other estimators is similar to that found in Table 7, reasserting the differences between estimation techniques. Note that we have removed any insignificant dynamics. Across sectors the estimate of the autoregressive parameter lies around 0.5 implying moderate persistence for the process of the dependent variable. Regarding the indicators, specialization and competition are influencing economic activity in all sectors albeit with different dynamics. Considering the immediate effects we find significant positive specialization effects and significant negative competition effects, while the lagged effects are vice versa. The results for the diversity

indicator are less pronounced, i.e. we find significant effects for one sector only.

In general the estimation results in Tables 7 and 8 suggest that dynamic externalities do matter for local economic activity in Morocco. Table 9 presents the long-term effects implied by the estimates of Table 8. The long-term effects of regional production and unit wage costs are elasticities, while those for the externality indicators are semi-elasticities. We find significant positive long-term effects for the specialization indicator in four out of five sectors. Regarding diversity we find a significant negative long-term effect in one sector only. For the competition indicator negative long-term effects are present in all sectors.

Summarizing the empirical evidence, in most sectors we do find positive specialization effects and negative diversity and competition effects. Note that the construction of the diversity indicator is such that a negative coefficient implies a positive effect on economic activity. Hence, we conclude that both specialization and diversity are beneficial for local economic development, while competition is harmful. As both Jacobs and Porter externalities agree that competition should stimulate economic activity, the evidence here is in favour of MAR externalities confirming the results of Bun and El Makhloufi (1999). However, in the present study we do find significant differences across sectors, so actually pooling is not vindicated.

5.4. Accuracy of the estimates

The samples analysed have the dimensions $T = 11$ and $N = \{18, 18, 29, 20, 10\}$. Hence, the simulation results of the previous section can shed some light on the differences in estimation results. We will consider only the accuracy of coefficient estimators. Considering the case $N = 10, T = 10$ the simulations show that the bias in the coefficient estimators is small for the AHL estimator, moderate for bias corrected LSDV versions and substantial negative bias is found for ordinary LSDV and GMM estimators. Increasing the number of cross-section observations to $N = 20$ leads to an improvement in the accuracy of all estimators, but still ordinary LSDV and GMM estimators have a substantial negative bias in estimating the autoregressive parameter. Note that the AHL estimator is almost unbiased, but its variance is at least twice the variance of the LSDV estimator.

With these simulation results in mind and analyzing the estimation results in Table 7 a remarkable resemblance is found between the various estimates of the autoregressive parameter and the picture showed by the simulations. Therefore, the ordinary LSDV and GMM results on γ are most likely biased downward. On the other hand, bias corrected LSDV variants seem to be much less biased downward. Because of its large dispersion, the AHL estimator sometimes fails completely in producing sensible and accurate estimates. This unstable behavior

emerges too in the simulations, where in a considerable number of cases the AHL estimate of the autoregressive parameter exceeded unity. Hence, it seems wise to use GMM rather than AHL estimates in the construction of bias corrected LSDV estimators, although the simulations did not show serious differences in performance.

6. Concluding remarks

In this study the finite sample properties of various IV and GMM coefficient estimators for dynamic panel data models have been compared through Monte Carlo experiments with the inconsistent LSDV estimator and corrected consistent versions of LSDV. Kiviet (1995) compared in a similar Monte Carlo study the bias and efficiency of these estimators in panel data sets with T small and N moderately large and concluded that there is not a superior technique over a broad range of parameter values for this model. However, in the present study we find that when both T and N are rather small the bias corrected LSDV estimator has almost uniformly lower mean squared error in comparison to IV and GMM methods. This result is in line with results of Judson and Owen (1999) who used a slightly different design regarding the parameter values and sample dimensions. Also results of Harris and Mátyás (1996), who used a totally different design, coincide with the findings in this study.

Adding an extra term to the approximation formula for the bias of the LSDV estimator, as derived in Kiviet (1999), does not improve the finite sample performance of this estimator notably. In fact, the present simulation results indicate that including $O(T^{-1})$ and $O(T^{-2})$ terms and omitting the $O(N^{-1}T^{-1})$ contributions accounts for most of the bias in the LSDV estimator. Also the sensitivity of the corrected LSDV estimator to the choice of the preliminary consistent estimator (AHL or GMM) is found to be low, although the finite sample performance of AHL is rather different as compared with GMM.

As far as the variance estimators are concerned, we find that for the corrected LSDV estimator variance estimation is highly nonstandard when T is finite. For a particular implementation of corrected LSDV, i.e. when the estimate of the bias approximation is based on the AHL estimator, an analytical expression of its asymptotic variance has been found. In addition, bootstrap consistent variance estimators of LSDVc and LSDVc^{AHL} have been examined. The simulation results show that the analytical variance expression performs poorly for high values of the autoregressive parameter, which can be explained by the unstable behavior of the AHL estimator for these parameter values. On the other hand, the bootstrap variance estimator is relatively accurate in many cases. Also it is found that biases in LSDV and GMM variance estimators can be considerable, especially when T is

small as is the case in the typical panel data set. The anomalies of these methods in small samples are even more evident when the empirical null distributions of ordinary t statistics are considered. The true rejection probabilities of asymptotic t tests can differ substantially from the nominal size, which leads in many cases to incorrectly rejecting the null hypothesis either too often or far too little.

The available simulation evidence becomes helpful when analyzing the relationship between dynamic externalities and local economic activity in Morocco. We find some differences in estimation results when analyzing five one-digit industrial sectors in isolation. In general, the sector specific results show that both specialization and diversity are beneficial for local economic development, while competition is harmful. As both Jacobs and Porter externalities agree that competition should stimulate economic activity, the evidence here is in favour of MAR externalities confirming the results of Bun and El Makhloufi (1999) but revealing significant heterogeneity at the same time. As the available samples have both T and N relatively small, the simulation results can clarify the differences between estimators. The pattern of the coefficient estimates is remarkably in line with the corresponding simulation results. The anomalies of some estimation techniques are clearly visible and the superior performance of the corrected LSDV coefficient estimator in finite samples with both T and N small is apparent here.

References

[1] Ahn, S.C. and P. Schmidt (1995), Efficient estimation of models for dynamic panel data, *Journal of Econometrics* 68, 5-27.

[2] Alvarez, J., and M. Arellano (1998), The time series and cross-section asymptotics of dynamic panel data estimators, *CEMFI working paper* no. 9808, Madrid.

[3] Anderson, T.W. and C. Hsiao (1982), Formulation and Estimation of Dynamic Models using Panel Data, *Journal of Econometrics* 18, 47-82.

[4] 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.

[5] Bun, M.J.G., and A. El Makhloufi (1999), Dynamic externalities and regional economic activity: panel data evidence for Morocco, University of Amsterdam, mimeo.

[6] De Lucio, J.J., Herce J.J., and A. Goicolea (1998), The effects of externalities on value added and productivity growth in Spanish industry, *Documento de trabajo* 98-05, FEDEA and Universidad Complutense of Madrid.

[7] Glaeser, E.L., Kallal, H.D., Scheinkman, J.A. and A. Shleifer (1992), Growth in cities, *Journal of Political Economy* 100, 61, 1127-1152.

[8] Harris, M.N., and L. Mátyás (1996), A comparative analysis of different estimators for dynamic panel data models, Monash University, Australia.

[9] Henderson V., Kuncoro. A. and Turner M. (1995), Industrial development in cities, *Journal of Political Economy* 103, 51, 1067-85.

[10] Jacobs, J. (1969), *The economy of cities*, Vintage, New York.

[11] Judson, R.A. and A.L. Owen (1999), Estimating dynamic panel data models: a guide for macroeconomists, *Economics Letters* 65, 9-15.

[12] Kiviet, J.F. (1995), On bias, inconsistency, and efficiency of various estimators in dynamic panel data models, *Journal of Econometrics* 68, 53-78.

[13] Kiviet, J.F. (1999), Expectations of expansions for estimators in a dynamic panel data model; some results for weakly exogenous regressors, in C. Hsiao, K. Lahiri, L-F Lee and M.H. Pesaran (eds.), *Analysis of Panels and Limited Dependent Variables*, Cambridge University Press, Cambridge.

[14] McLeod, A.I., and K.W. Hipel (1978), Simulation procedures for Box-Jenkins models, *Water Resources Research* 14, 969-975.

[15] Nickell, S. (1981), Biases in dynamic models with fixed effects, *Econometrica* 49, 1417-1426.

[16] Porter, M.E. (1990), *The competitive advantage of nations*, Free Press, New York.

[17] Romer, P.M. (1986), Increasing returns and long-run growth, *Journal of political economy* 94, 1002-1037.

[18] Sevestre, P., and A. Trognon (1985), A note on autoregressive error component models, *Journal of Econometrics* 28, 231-245.

[19] Sevestre, P., and A. Trognon (1996), Dynamic linear models. In L. Mátyás and P. Sevestre (eds.), The Econometrics of Panel Data, Kluwer Academic Publishers, The Netherlands.

A. The limiting distribution of LSDVc

First we derive the limiting distribution of $\sqrt{N}(\hat{\delta}_h^* - \delta^*)$, the second term of (3.29), with δ^* given in (3.23) and $\hat{\delta}_h^*$ given in (3.27). Next, we combine both terms in (3.29) and show that the resulting estimator is a sum of three zero-mean vectors. A central limit theorem applies and the corrected LSDV estimator is found to have a limiting normal distribution. The derivation of the asymptotic covariance matrix is pursued only for the case that the AHL estimator is used as a preliminary consistent estimator. Because of the complexity of the dependence of the instruments on ε for the GMM estimator we have not derived the expression for the variance in that case.

The expression (3.27) for $\hat{\delta}_h^*$ contains a factor

$$\hat{\sigma}_{\varepsilon,h}^2 \left(\frac{1}{1-\hat{\gamma}_h} - \frac{1-\hat{\gamma}_h^T}{T\left(1-\hat{\gamma}_h\right)^2} \right), \quad h = IV,\ GMM \tag{A.1}$$

whose components can be approximated as follows. First we have

$$\begin{aligned}
\hat{\sigma}_{\varepsilon,h}^2 &= \sigma_\varepsilon^2 + \left(\hat{\sigma}_{\varepsilon,h}^2 - \sigma_\varepsilon^2\right) \\
&= \sigma_\varepsilon^2 + \left(\frac{(y - W\hat{\delta}_h)' A (y - W\hat{\delta}_h)}{N(T-1)-(K+1)} - \sigma_\varepsilon^2 \right) \\
&= \sigma_\varepsilon^2 + \left(\frac{[\varepsilon - W(\hat{\delta}_h - \delta)]' A [\varepsilon - W(\hat{\delta}_h - \delta)]}{N(T-1)-(K+1)} - \sigma_\varepsilon^2 \right) \\
&= \sigma_\varepsilon^2 + \left(\frac{\varepsilon' A \varepsilon}{N(T-1)-(K+1)} - \sigma_\varepsilon^2 \right) + O_p(N^{-1}),
\end{aligned}$$

where the second term is $O_p(N^{-1/2})$. This follows from

$$E\left(\frac{\varepsilon' A \varepsilon}{N(T-1)-(K+1)} - \sigma_\varepsilon^2 \right) = \sigma_\varepsilon^2 \left(\frac{N\,(T-1)}{N(T-1)-(K+1)} - 1 \right) = O(N^{-1})$$

and

$$\mathrm{Var}\left(\frac{\varepsilon' A \varepsilon}{N(T-1)-(K+1)} - \sigma_\varepsilon^2 \right) = O(N^{-1}),$$

which results from $\mathrm{Var}(\varepsilon' A \varepsilon) = 2\sigma_\varepsilon^4 tr(AA) = 2N(T-1)\sigma_\varepsilon^4$. The order of the remainder term can be established from $W'A\varepsilon = O_p(N^{1/2})$ and $(\hat{\delta}_h - \delta) = O_p(N^{-1/2})$. Given the magnitude of the remainder term we can simplify the second term using

$$\left(\frac{\varepsilon' A \varepsilon}{N(T-1)-(K+1)} - \sigma_\varepsilon^2 \right) = \left(\frac{\varepsilon' A \varepsilon}{N(T-1)} - \sigma_\varepsilon^2 \right) + O_p(N^{-1}),$$

so we have

$$\hat{\sigma}^2_{\varepsilon,h} = \sigma^2_\varepsilon + \left(\frac{\varepsilon' A \varepsilon}{N(T-1)} - \sigma^2_\varepsilon \right) + O_p(N^{-1}). \tag{A.2}$$

A next component of (A.1) is

$$\begin{aligned} \frac{1}{1-\hat{\gamma}_h} &= \frac{1}{1-\gamma-(\hat{\gamma}_h-\gamma)} \\ &= \frac{1}{1-\gamma}\left[1-\frac{1}{1-\gamma}(\hat{\gamma}_h-\gamma)\right]^{-1} \\ &= \frac{1}{1-\gamma}\left[1+\frac{1}{1-\gamma}(\hat{\gamma}_h-\gamma)\right]+O_p(N^{-1}) \\ &= \frac{1}{1-\gamma}+\left(\frac{1}{1-\gamma}\right)^2(\hat{\gamma}_h-\gamma)+O_p(N^{-1}), \end{aligned} \tag{A.3}$$

where the second term is $O_p(N^{-1/2})$. From (A.3) we easily find

$$\left(\frac{1}{1-\hat{\gamma}_h}\right)^2 = \left(\frac{1}{1-\gamma}\right)^2 + 2\left(\frac{1}{1-\gamma}\right)^3(\hat{\gamma}_h-\gamma)+O_p(N^{-1}). \tag{A.4}$$

Next we obtain

$$\begin{aligned} 1-\hat{\gamma}_h^T &= 1-[\gamma+(\hat{\gamma}_h-\gamma)]^T \\ &= 1-\gamma^T-T\gamma^{T-1}(\hat{\gamma}_h-\gamma)+O_p(N^{-1}). \end{aligned} \tag{A.5}$$

For the second factor of (A.1) the results (A.3) - (A.5) yield

$$\begin{aligned} \frac{1}{1-\hat{\gamma}_h}-\frac{1-\hat{\gamma}_h^T}{T(1-\hat{\gamma}_h)^2} &= \frac{1}{1-\gamma}-\frac{1-\gamma^T}{T(1-\gamma)^2} \\ &\quad +\left[\frac{1+\gamma^{T-1}}{(1-\gamma)^2}-\frac{2}{T}\frac{1-\gamma^T}{(1-\gamma)^3}\right](\hat{\gamma}_h-\gamma) \\ &\quad +O_p(N^{-1}). \end{aligned} \tag{A.6}$$

>From (A.2) and (A.6) we obtain:

$$\begin{aligned} &\hat{\sigma}^2_{\varepsilon,h}\left(\frac{1}{1-\hat{\gamma}_h}-\frac{1-\hat{\gamma}_h^T}{T(1-\hat{\gamma}_h)^2}\right)-\sigma^2_\varepsilon\left(\frac{1}{1-\gamma}-\frac{1-\gamma^T}{T(1-\gamma)^2}\right)= \\ &\quad \sigma^2_\varepsilon\left[\frac{1+\gamma^{T-1}}{(1-\gamma)^2}-\frac{2}{T}\frac{1-\gamma^T}{(1-\gamma)^3}\right](\hat{\gamma}_h-\gamma) \\ &\quad +\left(\frac{1}{1-\gamma}-\frac{1-\gamma^T}{T(1-\gamma)^2}\right)\left(\frac{\varepsilon' A\varepsilon}{N(T-1)}-\sigma^2_\varepsilon\right) \\ &\quad +O_p(N^{-1}) \end{aligned}$$

giving

$$\begin{aligned}
& \sqrt{N}\left(\hat{\delta}_h^* - \delta^*\right) \qquad (A.7)\\
= \; & -\sigma_\varepsilon^2\left[\frac{1+\gamma^{T-1}}{(1-\gamma)^2} - \frac{2}{T}\frac{1-\gamma^T}{(1-\gamma)^3}\right]\Sigma_{WAW}^{-1}e_1e_1^{'}\sqrt{N}\left(\hat{\delta}_h - \delta\right)\\
& -tr(\Pi_T)\Sigma_{WAW}^{-1}e_1\sqrt{N}\left(\frac{\varepsilon' A\varepsilon}{N(T-1)} - \sigma_\varepsilon^2\right) + O_p(N^{-1/2})\\
= \; & -\sigma_\varepsilon^2\bar{d}_1e_1'\sqrt{N}\left(\hat{\delta}_h - \delta\right) - \bar{d}_2\sqrt{N}\left(\frac{\varepsilon' A\varepsilon}{N(T-1)} - \sigma_\varepsilon^2\right) + O_p(N^{-1/2}),
\end{aligned}$$

where $\bar{d}_1$ and $\bar{d}_2$ have been defined implicitly.

Now we find from (3.29) that the limiting distribution of $\sqrt{N}(\hat{\delta}_{LSDVc,h} - \delta)$ must be the same as the limiting distribution of the sum of the three zero-mean vectors v_0, $v_{1,h}$ and v_2, where

$$\begin{aligned}
v_0 &= \Sigma_{WAW}^{-1}\left(\frac{1}{\sqrt{N}}[\bar{W}'A\varepsilon + \left(\varepsilon'\Pi\varepsilon - \sigma_\varepsilon^2 tr(\Pi)\right)e_1]\right) \qquad (A.8)\\
v_{1,IV} &= \sigma_\varepsilon^2\bar{d}_1e_1'\Sigma_{ZW^*}^{-1}\left(\frac{1}{\sqrt{N}}Z'D\varepsilon\right)\\
v_{1,GMM} &= \sigma_\varepsilon^2\bar{d}_1e_1'(\Sigma_{Z^*W^*}'\Sigma_{Z^*HZ^*}^{-1}\Sigma_{Z^*W^*})^{-1}\Sigma_{Z^*W^*}'\Sigma_{Z^*HZ^*}^{-1}\left(\frac{1}{\sqrt{N}}Z^{*\prime}D\varepsilon\right)\\
v_2 &= \bar{d}_2\sqrt{N}\left(\frac{\varepsilon' A\varepsilon}{N(T-1)} - \sigma_\varepsilon^2\right).
\end{aligned}$$

Hence, the limiting distribution of the corrected LSDV estimator is

$$\sqrt{N}(\hat{\delta}_{LSDVc,h} - \delta) \xrightarrow[N\to\infty]{d} \mathcal{N}\left[0,\ \Omega_{LSDVc,h}\right]$$

where

$$\Omega_{LSDVc,h} = \lim_{N\to\infty}\mathrm{Var}(v_0 + v_{1,h} + v_2). \qquad (A.9)$$

Taking now the AHL estimator $\hat{\delta}_{IV}$ for $\hat{\delta}_h$, some of the components of $\Omega_{LSDVc,IV}$ have been established before, so we easily find for $N \to \infty$

$$\begin{aligned}
\mathrm{Var}(v_0) &= \sigma_\varepsilon^2\Sigma_{WAW}^{-1} + \sigma_\varepsilon^4 tr(\Pi_T\Pi_T)\Sigma_{WAW}^{-1}e_1e_1'\Sigma_{WAW}^{-1}\\
\mathrm{Var}(v_{1,IV}) &= \sigma_\varepsilon^6 e_1'\Sigma_{ZW^*}^{-1}\Sigma_{ZHZ}(\Sigma_{ZW^*}')^{-1}e_1\bar{d}_1\bar{d}_1'\\
\mathrm{Var}(v_2) &= \frac{2\sigma_\varepsilon^4}{T-1}\bar{d}_2\bar{d}_2',
\end{aligned}$$

where we used results of (3.25), (3.5) and one mentioned above (A.2) respectively. Next we obtain for $N \rightarrow \infty$

$$\begin{aligned} E(v_0 v_2') &= \Sigma_{WAW}^{-1} E[\bar{W}' A\varepsilon + \left(\varepsilon'\Pi\varepsilon - \sigma_\varepsilon^2 tr(\Pi)\right)] e_1 \left(\frac{\varepsilon' A\varepsilon}{N(T-1)} - \sigma_\varepsilon^2\right) \bar{d}_2' \\ &= \Sigma_{WAW}^{-1} e_1 \bar{d}_2' E\left(\varepsilon'\Pi\varepsilon - \sigma_\varepsilon^2 tr(\Pi)\right) \left(\frac{\varepsilon' A\varepsilon}{N(T-1)} - \sigma_\varepsilon^2\right) \\ &= 2\sigma_\varepsilon^4 \frac{tr(\Pi_T)}{T-1} \Sigma_{WAW}^{-1} e_1 \bar{d}_2' \\ &= \frac{2\sigma_\varepsilon^4}{T-1} \bar{d}_2 \bar{d}_2'. \end{aligned}$$

To derive covariances with $v_{1,IV}$ we have to express the stochastic nature of the instruments $Z = [y_{-2} \vdots DX]$ and the disturbances $D\varepsilon$ explicitly. We may write

$$Z - E(Z \mid X, \eta) = KL\Gamma\varepsilon e_1',$$

see Kiviet (1999), where the $N(T-1) \times NT$ matrix K performs lag operations. More precisely, $K = I_N \otimes K_T$ with the $(T-1) \times T$ matrix K_T defined as

$$K_T = \begin{bmatrix} 1 & 0 & \cdot & \cdot & \cdot & 0 \\ 0 & 1 & 0 & \cdot & & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & & \cdot & 1 & 0 & \cdot \\ 0 & \cdot & \cdot & 0 & 1 & 0 \end{bmatrix}.$$

This can be used to obtain

$$\begin{aligned} E(v_0 v_{1,IV}') &= \frac{\sigma_\varepsilon^2}{N} \Sigma_{WAW}^{-1} E\{[\bar{W}' A\varepsilon + \left(\varepsilon'\Pi\varepsilon - \sigma_\varepsilon^2 tr(\Pi_T)\right) e_1] \\ &\qquad \times (\varepsilon' D'\bar{Z} + \varepsilon' D' KL\Gamma\varepsilon e_1')\} (\Sigma_{ZW^*}')^{-1} e_1 \bar{d}_1' \\ &= \frac{\sigma_\varepsilon^2}{N} \Sigma_{WAW}^{-1} E(\bar{W}' A\varepsilon\varepsilon' D'\bar{Z}) (\Sigma_{ZW^*}')^{-1} e_1 \bar{d}_1' \\ &\quad + \frac{\sigma_\varepsilon^2}{N} \Sigma_{WAW}^{-1} E\{\left(\varepsilon'\Pi\varepsilon - \sigma_\varepsilon^2 tr(\Pi)\right) \varepsilon' D' KL\Gamma\varepsilon e_1 e_1'\} (\Sigma_{ZW^*}')^{-1} e_1 \bar{d}_1' \\ &= \frac{\sigma_\varepsilon^4}{N} \Sigma_{WAW}^{-1} \bar{W}' AD'\bar{Z} (\Sigma_{ZW^*}')^{-1} e_1 \bar{d}_1' + \\ &\quad \frac{\sigma_\varepsilon^6}{N} \{tr(\Pi D' KL\Gamma) + tr(\Pi\Gamma' L' K' D)\} \Sigma_{WAW}^{-1} (\Sigma_{ZW^*}')^{-1} e_1 \bar{d}_1' \end{aligned}$$

and

$$E(v_1 v_2') = \sigma_\varepsilon^2 \bar{d}_1 e_1' \Sigma_{ZW^*}^{-1} e_1 E\left(\varepsilon'\Gamma' L' K' D\varepsilon\right) \left(\frac{\varepsilon' A\varepsilon}{N(T-1)} - \sigma_\varepsilon^2\right) \bar{d}_2'$$

$$
\begin{aligned}
&= 2\sigma_\varepsilon^6 tr(D'KL\Gamma A)\bar{d}_1 e_1' \Sigma_{ZW^*}^{-1} e_1 \\
&= O,
\end{aligned}
$$

noting that $tr(D'KL\Gamma A) = 0$. Defining

$$
\begin{aligned}
\Sigma_{WADZ} &= \operatorname*{plim}_{N\to\infty}\left(\frac{1}{N}W'AD'Z\right) \\
&= \lim_{N\to\infty}\left(\frac{1}{N}\bar{W}'AD'\bar{Z}\right) + \sigma_\varepsilon^2 tr(\Pi_T' D_T' K_T L_T \Gamma_T) e_1 e_1',
\end{aligned}
$$

we find

$$
\begin{aligned}
\lim_{N\to\infty} E(v_0 v_{1,IV}') &= \sigma_\varepsilon^4 \Sigma_{WAW}^{-1} \Sigma_{WADZ} (\Sigma_{ZW^*}')^{-1} e_1 \bar{d}_1' \\
&\quad + \sigma_\varepsilon^6 tr(\Pi_T \Gamma_T' L_T' K_T' D_T) \Sigma_{WAW}^{-1} (\Sigma_{ZW^*}')^{-1} e_1 \bar{d}_1'.
\end{aligned}
$$

Collecting terms this yields

$$
\begin{aligned}
\Omega_{LSDVc,IV} &= \sigma_\varepsilon^2 \Sigma_{WAW}^{-1} + \sigma_\varepsilon^4 tr(\Pi_T \Pi_T) \Sigma_{WAW}^{-1} e_1 e_1' \Sigma_{WAW}^{-1} \\
&\quad + \sigma_\varepsilon^6 e_1' \Sigma_{ZW^*}^{-1} \Sigma_{ZHZ} (\Sigma_{ZW^*}')^{-1} e_1 \bar{d}_1 \bar{d}_1' + (6/(T-1))\, \sigma_\varepsilon^4 \bar{d}_2 \bar{d}_2' \\
&\quad + \sigma_\varepsilon^4 [\Sigma_{WAW}^{-1} \Sigma_{WADZ} (\Sigma_{ZW^*}')^{-1} e_1 \bar{d}_1' + \bar{d}_1 e_1' (\Sigma_{ZW^*})^{-1} \Sigma_{WADZ}' \Sigma_{WAW}^{-1}] \\
&\quad + \sigma_\varepsilon^6 tr(\Pi_T \Gamma_T' L_T' K_T' D_T) [\Sigma_{WAW}^{-1} (\Sigma_{ZW^*}')^{-1} e_1 \bar{d}_1' + \bar{d}_1 e_1' (\Sigma_{ZW^*})^{-1} \Sigma_{WAW}^{-1}].
\end{aligned} \tag{A.10}
$$

For the case where the $\sqrt{N}$-consistent estimator $\hat{\delta}_{GMM}$ of (3.6) is employed, we find that $\sqrt{N}\left(\hat{\delta}_{LSDVc,GMM} - \delta\right)$ has a limiting distribution where the variance $V_{LSDVc,GMM}$ can be obtained in the same way. The only difference is that we then have

$$
v_{1,GMM} = \sigma_\varepsilon^2 \bar{d}_1 e_1' (\Sigma_{Z^*W^*}' \Sigma_{Z^*HZ^*}^{-1} \Sigma_{Z^*W^*})^{-1} \Sigma_{Z^*W^*}' \Sigma_{Z^*HZ^*}^{-1} \left(\frac{1}{\sqrt{N}} Z^{*\prime} \Delta\varepsilon\right).
$$

Because of the complexity of the dependence of the instruments Z^* on ε we have not derived the expression for the variance $V_{LSDVc,GMM}$, but it is feasible.

B. Details on the simulation design

The simulation design is basically the same as in Kiviet (1995), especially for the case where all variables are stationary. The generating equation for y_{it} is

$$
y_{it} = \gamma y_{i,t-1} + \beta' x_{it} + \eta_i + \varepsilon_{it} \tag{B.1}
$$

for individuals $i = 1, ..., N$ and time periods $t = 1, ..., T$. A latent variable v_{it}, free of the individual specific effect in y_{it}, is introduced:

$$v_{it} = y_{it} - \frac{1}{1-\gamma}\eta_i. \tag{B.2}$$

Substitution of (B.2) in (B.1) leads to

$$v_{it} = \gamma v_{i,t-1} + \beta' x_{it} + \varepsilon_{it}. \tag{B.3}$$

For $K = 1$ and upon omitting the index i, the model for every cross-section unit becomes

$$\begin{aligned} v_t &= \gamma v_{t-1} + \beta x_t + \varepsilon_t \\ x_t &= \rho x_{t-1} + \xi_t. \end{aligned} \tag{B.4}$$

Using L for the lag-operator and substituting the second equation of (B.4) into the first, we obtain for $|\rho| < 1$

$$\begin{aligned} v_t &= \beta \frac{1}{(1-\gamma L)(1-\rho L)}\xi_t + \frac{1}{1-\gamma L}\varepsilon_t \\ &= \beta\phi_t + \psi_t. \end{aligned} \tag{B.5}$$

The latent variable v_t consists of two independent components, viz. $\beta\phi_t$ and ψ_t, which are an AR(2) and AR(1) process respectively. Data for the processes ξ_t and ε_t are obtained through sampling independently from $\mathcal{IIN}(0, \sigma_\xi^2)$ and $\mathcal{IIN}(0, \sigma_\varepsilon^2)$. Next σ_ξ^2 is determined through fixing the signal-noise ratio, which is a measure of the explanatory power of the regressors, defined as $\mathrm{Var}(v_{it} - \varepsilon_{it})/\mathrm{Var}(\varepsilon_{it})$. Scaling with respect to σ_ε^2, i.e. $\sigma_\varepsilon = 1$, this is in the stationary case

$$\begin{aligned} \sigma_s^2 &= \mathrm{Var}(v_{it} - \varepsilon_{it}) \\ &= \beta^2\sigma_\xi^2[1 + \frac{(\gamma+\rho)^2}{1+\gamma\rho}(\gamma\rho - 1) - (\gamma\rho)^2]^{-1} + \frac{\gamma^2}{1-\gamma^2}. \end{aligned} \tag{B.6}$$

When values for γ, β, ρ and for σ_s^2 have been chosen, σ_ξ^2 is determined through

$$\sigma_\xi^2 = \frac{1}{\beta^2}[\sigma_s^2 - \frac{\gamma^2}{1-\gamma^2}][1 + \frac{(\gamma+\rho)^2}{1+\gamma\rho}(\gamma\rho - 1) - (\gamma\rho)^2]. \tag{B.7}$$

The series ϕ_t, ψ_t and x_t, including starting values ϕ_0, ϕ_1, ψ_0 and x_0, are obtained by employing a procedure described in McLeod and Hipel (1978). Let z_t be a normal stationary $AR(p)$ process, i.e.

$$z_t = \alpha_1 z_{t-1} + ... + \alpha_p z_{t-p} + u_t, \ t = 1, ..., T \tag{B.8}$$

with $u_t \sim \mathcal{IIN}(0, \sigma_u^2)$. It is easy to generate a series $u_1, ..., u_T$. Consider now the first p observations $z = (z_1, ..., z_p)$ and $u = (u_1, ..., u_p)$. Let $\sigma_u^2 \Sigma$ be the covariance matrix of z and consider the Cholesky decomposition of Σ,

$$\Sigma = MM', \tag{B.9}$$

where the matrix M is lower triangular. It is easily seen that z and Mu have the same distribution, and hence a p element vector z can be generated easily from the first p elements of u upon making use of the autocovariance function of the $AR(p)$ process (which determines Σ) and the Cholesky decomposition M of the covariance matrix Σ. The remaining $T - p$ observations $t = p + 1, ..., T$ can be generated now recursively according to (B.8).

Hence, once starting values have been obtained, series x and y can be generated according to

$$\begin{aligned} \phi_t &= (\gamma + \rho)\phi_{t-1} - \gamma\rho\phi_{t-2} + \xi_t \\ \psi_t &= \gamma\psi_{t-1} + \varepsilon_t \\ x_t &= \rho x_{t-1} + \xi_t \\ y_t &= \beta\phi_t + \psi_t + \frac{1}{1-\gamma}\eta, \end{aligned} \tag{B.10}$$

where the vector of individual specific effects η is drawn from $\mathcal{IIN}(0, \sigma_\eta^2)$. The standard deviation σ_η of the individual specific effect is determined by

$$\sigma_\eta = \mu\sigma_\varepsilon(1 - \gamma). \tag{B.11}$$

In this way the impact on y of both the individual specific effect η and the general disturbance term ε are independent of the values of γ and β, and the relative importance of both shocks can be varied through choosing different values for μ. For example, when μ is unity the impact on y of both shocks is the same.

A different situation arises when the exogenous variable x is non-stationary. In this case model (B.4) for the latent variable v_t becomes

$$\begin{aligned} v_t &= \gamma v_{t-1} + \beta x_t + \varepsilon_t \\ x_t &= x_{t-1} + \xi_t. \end{aligned} \tag{B.12}$$

The exogenous variable can be written as

$$x_t = x_0 + \sum_{s=1}^{t} \xi_s. \tag{B.13}$$

Now a similar expression as in (B.5) can be obtained through substitution of (B.13) into the equation for v_t, i.e.

$$\begin{aligned} v_t &= \frac{\beta}{1-\gamma}x_0 + \beta\sum_{s=1}^{t}(1-\gamma L)^{-1}\xi_s + (1-\gamma L)^{-1}\varepsilon_t \\ &= \frac{\beta}{1-\gamma}x_0 + \beta\sum_{s=1}^{t}\phi_s^* + \psi_t. \end{aligned} \tag{B.14}$$

The latent variable has again two random components, viz. the $AR(1)$ process ψ_t and the partial sum of an $AR(1)$ process. Starting values for the processes ϕ_t^* and ψ_t can be obtained again by using the procedure described above. Drawings for y and x can now be generated from

$$\begin{aligned} \phi_t^* &= \gamma\phi_{t-1}^* + \xi_t \\ \psi_t &= \gamma\psi_{t-1} + \varepsilon_t \\ x_t &= x_{t-1} + \xi_t \\ y_t &= \frac{\beta}{1-\gamma}x_0 + \beta\sum_{s=1}^{t}\phi_s^* + \psi_t + \frac{1}{1-\gamma}\eta, \end{aligned} \tag{B.15}$$

after choosing a value for x_0. This has been set equal to zero, without loss of generality.

While the generation of the data in case of a non-stationary exogenous variable does not cause any complications, the control over the signal is not so straightforward. The signal in the non-stationary case is not constant through time, but increasing. We find

$$\begin{aligned} \sigma_{s,t}^2 &= Var(v_{it} - \varepsilon_{it}) \\ &= \frac{\beta^2\sigma_\xi^2}{1-\gamma^2}[t(1+\frac{2\gamma}{1-\gamma}) - \frac{2\gamma(1-\gamma^t)}{(1-\gamma)^2}] + \frac{\gamma^2}{1-\gamma^2}\sigma_\varepsilon^2,\ t = 1, ..., T. \end{aligned} \tag{B.16}$$

To control the signal in the experiments the mean value of $\sigma_{s,t}^2$ (over $t = 1, ..., T$) has been fixed. In this way it was tried to make the outcomes for $\rho = 1$ more or less comparable with the results for the stationary case.

Table 1:
Bias, standard deviation and root mean-squared error of estimators for γ and β in case of stationary y and x

	#	*Bias* γ	*Bias* β	*Std* γ	*Std* β	*Rmse* γ	*Rmse* β
			$N = 10, T = 20$				
$LSDV$	0	-0.074	0.013	0.057	0.084	0.095	0.086
$LSDVc$	0	-0.016	-0.014	0.060	0.081	0.063	0.082
$LSDVc^{AHL}$	0	-0.009	-0.012	0.061	0.083	0.062	0.084
AHL	37	-0.006	-0.009	0.133	0.151	0.133	0.152
$GMMa$	4	-0.079	0.022	0.116	0.143	0.141	0.145
$GMMb$	0	-0.116	0.073	0.078	0.122	0.140	0.145
$GMMc$	0	-0.109	0.054	0.069	0.108	0.130	0.123
			$N = 10, T = 10$				
$LSDV$	0	-0.149	0.026	0.088	0.149	0.175	0.153
$LSDVc$	0	-0.043	-0.011	0.096	0.144	0.108	0.144
$LSDVc^{AHL}$	0	-0.024	-0.010	0.099	0.147	0.104	0.147
AHL	77	0.002	-0.016	0.242	0.215	0.242	0.215
$GMMa$	19	-0.098	0.010	0.187	0.201	0.213	0.202
$GMMb$	1	-0.166	0.038	0.132	0.184	0.214	0.190
$GMMc$	0	-0.165	0.039	0.121	0.177	0.206	0.183
			$N = 20, T = 10$				
$LSDV$	0	-0.144	0.031	0.062	0.099	0.158	0.106
$LSDVc$	0	-0.037	-0.009	0.069	0.096	0.080	0.096
$LSDVc^{AHL}$	0	-0.020	-0.010	0.070	0.098	0.074	0.099
AHL	49	-0.006	-0.004	0.151	0.158	0.152	0.158
$GMMa$	15	-0.059	0.008	0.135	0.153	0.148	0.153
$GMMb$	0	-0.111	0.026	0.101	0.140	0.151	0.144
$GMMc$	0	-0.115	0.023	0.095	0.134	0.150	0.137

* Figures are averages over three experiments, i.e. $\gamma = \{0.8, 0.5, 0.2\}$

* # is number of cases where the estimate of γ is equal or larger than 1 in absolute value

Table 2:
Bias, standard deviation and root mean-squared error of estimators for γ, and β in case of non-stationary y and x

	#	*Bias* γ	*Bias* β	*Std* γ	*Std* β	*Rmse* γ	*Rmse* β
			$N = 10, T = 20$				
$LSDV$	0	-0.082	0.045	0.061	0.164	0.103	0.171
$LSDVc$	0	-0.012	0.003	0.065	0.153	0.066	0.154
$LSDVc^{AHL}$	0	-0.004	0.000	0.065	0.159	0.066	0.159
AHL	36	-0.003	0.001	0.152	0.323	0.152	0.323
$GMMa$	3	-0.095	0.007	0.131	0.313	0.162	0.313
$GMMb$	0	-0.150	0.019	0.095	0.280	0.178	0.282
$GMMc$	0	-0.131	0.037	0.079	0.230	0.153	0.235
			$N = 10, T = 10$				
$LSDV$	0	-0.154	0.090	0.089	0.200	0.180	0.220
$LSDVc$	0	-0.041	0.021	0.099	0.191	0.109	0.193
$LSDVc^{AHL}$	0	-0.025	0.026	0.105	0.205	0.110	0.207
AHL	143	0.009	0.008	0.838	0.455	0.839	0.455
$GMMa$	17	-0.170	-0.012	0.245	0.297	0.298	0.297
$GMMb$	0	-0.213	0.022	0.149	0.268	0.260	0.270
$GMMc$	0	-0.191	0.045	0.132	0.256	0.233	0.261
			$N = 20, T = 10$				
$LSDV$	0	-0.145	0.091	0.062	0.142	0.159	0.169
$LSDVc$	0	-0.033	0.024	0.068	0.135	0.078	0.137
$LSDVc^{AHL}$	0	-0.014	0.020	0.072	0.140	0.075	0.142
AHL	82	0.004	0.001	0.299	0.230	0.299	0.230
$GMMa$	14	-0.118	-0.006	0.195	0.218	0.228	0.218
$GMMb$	0	-0.166	0.012	0.130	0.204	0.211	0.205
$GMMc$	0	-0.153	0.028	0.116	0.196	0.193	0.199

* Figures are averages over three experiments, i.e. $\gamma = \{0.8, 0.5, 0.2\}$

* # is number of cases where the estimate of γ is equal or larger than 1 in absolute value

Table 3:
LSDV bias for stationary y and x

	$Bias\ \gamma$	$Bias\ \beta$	$Bias\ \gamma$	$Bias\ \beta$	$Bias\ \gamma$	$Bias\ \beta$
	$N = 10,\ T = 20$		$N = 10,\ T = 10$		$N = 20,\ T = 10$	
			$\gamma = 0.8,\ \rho = 0.8$			
$total$	-0.115	0.001	-0.232	0.005	-0.224	0.015
$B^{(0)}_{LSDV}$	-0.134	0.028	-0.388	0.032	-0.387	0.047
$B^{(1)}_{LSDV}$	-0.101	0.021	-0.215	0.018	-0.214	0.026
$B^{(2)}_{LSDV}$	-0.108	0.024	-0.218	0.020	-0.216	0.027
$B^{(3)}_{LSDV}$	-0.113	0.025	-0.232	0.021	-0.224	0.028
			$\gamma = 0.5,\ \rho = 0.8$			
$total$	-0.062	0.023	-0.126	0.039	-0.123	0.043
$B^{(0)}_{LSDV}$	-0.055	0.031	-0.136	0.054	-0.138	0.063
$B^{(1)}_{LSDV}$	-0.050	0.028	-0.109	0.043	-0.111	0.050
$B^{(2)}_{LSDV}$	-0.055	0.032	-0.115	0.048	-0.114	0.052
$B^{(3)}_{LSDV}$	-0.056	0.032	-0.120	0.050	-0.116	0.053
			$\gamma = 0.2,\ \rho = 0.8$			
$total$	-0.046	0.014	-0.088	0.035	-0.085	0.034
$B^{(0)}_{LSDV}$	-0.037	0.026	-0.084	0.049	-0.085	0.052
$B^{(1)}_{LSDV}$	-0.035	0.025	-0.073	0.043	-0.074	0.046
$B^{(2)}_{LSDV}$	-0.039	0.027	-0.079	0.047	-0.077	0.048
$B^{(3)}_{LSDV}$	-0.039	0.028	-0.080	0.048	-0.078	0.048

* $B^{(j)}_{LSDV}$, j=0,...,3, is given in (2.17) and (2.18)

Table 4:
LSDV bias for non-stationary y and x

	$Bias\ \gamma$	$Bias\ \beta$	$Bias\ \gamma$	$Bias\ \beta$	$Bias\ \gamma$	$Bias\ \beta$
	$N = 10,\ T = 20$		$N = 10,\ T = 10$		$N = 20,\ T = 10$	
			$\gamma = 0.8,\ \rho = 1.0$			
$total$	-0.114	0.044	-0.229	0.104	-0.214	0.109
$B^{(0)}_{LSDV}$	-0.137	0.066	-0.386	0.185	-0.384	0.168
$B^{(1)}_{LSDV}$	-0.103	0.050	-0.214	0.102	-0.213	0.093
$B^{(2)}_{LSDV}$	-0.111	0.054	-0.219	0.105	-0.215	0.095
$B^{(3)}_{LSDV}$	-0.116	0.056	-0.233	0.111	-0.223	0.097
			$\gamma = 0.5,\ \rho = 1.0$			
$total$	-0.075	0.048	-0.135	0.090	-0.127	0.090
$B^{(0)}_{LSDV}$	-0.075	0.053	-0.154	0.109	-0.159	0.105
$B^{(1)}_{LSDV}$	-0.068	0.047	-0.123	0.087	-0.127	0.084
$B^{(2)}_{LSDV}$	-0.075	0.052	-0.131	0.093	-0.131	0.087
$B^{(3)}_{LSDV}$	-0.077	0.053	-0.135	0.096	-0.134	0.089
			$\gamma = 0.2,\ \rho = 1.0$			
$total$	-0.056	0.042	-0.098	0.074	-0.095	0.075
$B^{(0)}_{LSDV}$	-0.055	0.044	-0.102	0.083	-0.105	0.080
$B^{(1)}_{LSDV}$	-0.051	0.041	-0.089	0.072	-0.092	0.070
$B^{(2)}_{LSDV}$	-0.056	0.045	-0.095	0.078	-0.095	0.073
$B^{(3)}_{LSDV}$	-0.057	0.046	-0.097	0.080	-0.097	0.074

* $B^{(j)}_{LSDV}$, j=0,...,3, is given in (2.17) and (2.18)

Table 5:
Bias of coefficient estimators and estimators of standard errors ($\gamma = 0.5$ and $\rho = 0.8$)

$Bias\ of:$	γ	β	σ_ε	$Std\ \gamma$	$Std\ \beta$
			$N = 10, T = 20$		
$LSDV$	-0.062	0.023	-2.5	-8.0	-3.4
$LSDVc$	-0.012	-0.006	-2.3	-11.1	-3.0
$LSDVc^{AHL}$	-0.006	-0.008	-2.2	-12.0	-2.6
AHL	-0.008	-0.007	0.6	0.7	0.6
$GMMa$	-0.071	0.025	-1.0	1.5	1.2
$GMMb$	-0.106	0.078	-2.0	-3.5	2.7
$GMMc$	-0.098	0.062	-2.3	-6.0	-1.7
			$N = 10, T = 10$		
$LSDV$	-0.126	0.039	-5.9	-9.1	-10.0
$LSDVc$	-0.028	-0.005	-5.1	-15.4/NA/-5.6	-8.5/NA/-7.2
$LSDVc^{AHL}$	-0.006	-0.006	-4.6	-19.5/4.4/-5.8	-8.3/-6.0/-6.8
AHL	-0.001	-0.013	0.7	-1.0	2.9
$GMMa$	-0.084	0.016	-3.8	-3.6	1.8
$GMMb$	-0.145	0.051	-5.7	-13.0	-1.3
$GMMc$	-0.144	0.052	-5.9	-14.3	-3.6
			$N = 20, T = 10$		
$LSDV$	-0.123	0.043	-6.1	-8.3	-9.6
$LSDVc$	-0.025	-0.003	-5.3	-15.7/NA/-5.0	-8.7/NA/-5.8
$LSDVc^{AHL}$	-0.010	-0.007	-5.1	-18.7/3.5/-3.0	-8.3/-5.1/-5.4
AHL	-0.009	-0.004	-1.5	-0.4	-1.3
$GMMa$	-0.054	0.011	-4.8	-2.1	-3.1
$GMMb$	-0.097	0.036	-6.0	-7.9	-3.2
$GMMc$	-0.100	0.034	-6.2	-7.8	-4.1

* The bias in estimators of asymptotic standard errors is in percentage of the true value

* Multiple $\cdot/\cdot/\cdot$ figures are based on (3.21), (3.31) and (3.32) respectively

Table 6:
Estimated true size of nominal 5% one- and two-sided t tests ($\gamma = 0.5$ and $\rho = 0.8$)

H_1 :	$\gamma > \gamma^0$	$\gamma < \gamma^0$	$\gamma \neq \gamma^0$	$\beta > \beta^0$	$\beta < \beta^0$	$\beta \neq \beta^0$
			$N = 10, T = 20$			
$LSDV$	1	31	23	11	3	8
$LSDVc$	4	10	8	4	7	7
$LSDVc^{AHL}$	6	9	8	4	7	6
AHL	3	6	5	4	5	5
$GMMa$	1	16	9	8	2	6
$GMMb$	0	42	30	23	0	15
$GMMc$	0	45	34	21	1	14
			$N = 10, T = 10$			
$LSDV$	0	47	35	13	3	10
$LSDVc$	4/NA/3	13/NA/11	10/NA/7	7/NA/7	8/NA/8	8/NA/8
$LSDVc^{AHL}$	8/1/5	9/7/7	11/4/7	7/6/7	8/7/8	7/7/8
AHL	2	6	4	4	4	4
$GMMa$	1	13	9	7	3	4
$GMMb$	0	35	25	11	2	8
$GMMc$	0	41	30	12	3	9
			$N = 20, T = 10$			
$LSDV$	0	69	59	19	2	13
$LSDVc$	4/NA/3	14/NA/11	13/NA/8	6/NA/6	8/NA/7	7/NA/7
$LSDVc^{AHL}$	7/2/3	12/8/8	12/4/6	6/6/6	8/8/7	7/6/8
AHL	2	7	4	5	6	5
$GMMa$	1	15	8	6	4	6
$GMMb$	1	30	21	12	2	7
$GMMc$	0	34	23	11	3	7

* Multiple $\cdot/\cdot/\cdot$ figures are based on (3.21), (3.31) and (3.32) respectively

Table 7:
Estimation results of specification (5.4) for chemical and allied products (T=11, N=29)

	$LSDV$	$LSDVc$	$LSDVc^{AHL}$	AHL	$GMMa$	$GMMb$	$GMMc$
$\ln var_{i,t-1}$	0.29	0.44	0.52	0.72	0.44	0.29	0.34
	(0.05)	(0.06)	(0.08)	(0.37)	(0.17)	(0.10)	(0.09)
$\ln trp_{it}$	0.53	0.42	0.31	0.41	0.26	0.34	0.37
	(0.19)	(0.16)	(0.19)	(0.30)	(0.24)	(0.23)	(0.23)
$\ln trp_{i,t-1}$	-0.05	-0.05	-0.06	-0.18	0.02	0.10	0.01
	(0.18)	(0.16)	(0.17)	(0.33)	(0.24)	(0.23)	(0.22)
$\ln wcap_{it}$	0.75	0.64	0.58	0.77	0.74	0.73	0.78
	(0.07)	(0.06)	(0.06)	(0.11)	(0.09)	(0.09)	(0.09)
$\ln wcap_{i,t-1}$	-0.15	-0.16	-0.18	-0.42	-0.21	-0.18	-0.18
	(0.08)	(0.07)	(0.07)	(0.27)	(0.14)	(0.11)	(0.10)
sp_{it}	0.60	0.52	0.49	0.46	0.48	0.55	0.54
	(0.09)	(0.08)	(0.08)	(0.12)	(0.11)	(0.11)	(0.11)
$sp_{i,t-1}$	-0.12	-0.11	-0.10	-0.44	-0.29	-0.21	-0.27
	(0.09)	(0.07)	(0.08)	(0.23)	(0.14)	(0.12)	(0.11)
dv_{it}	0.06	0.03	0.00	0.46	0.40	0.43	0.35
	(0.13)	(0.12)	(0.13)	(0.22)	(0.19)	(0.19)	(0.18)
$dv_{i,t-1}$	-0.24	-0.23	-0.25	-0.40	-0.35	-0.33	-0.34
	(0.15)	(0.12)	(0.13)	(0.22)	(0.19)	(0.18)	(0.18)
cp_{it}	-0.15	-0.13	-0.11	-0.17	-0.16	-0.16	-0.16
	(0.02)	(0.02)	(0.02)	(0.03)	(0.03)	(0.03)	(0.03)
$cp_{i,t-1}$	0.06	0.06	0.07	0.08	0.04	0.03	0.04
	(0.02)	(0.02)	(0.02)	(0.06)	(0.04)	(0.03)	(0.03)

* Figures in parentheses are standard errors

Table 8:
LSDVc estimation results of specification (5.4) for the five one-digit sectors

	food and agricultural products	textile and leather	chemical and allied products	metal and machinery	electric and electronic manufacturing
$\ln var_{i,t-1}$	0.48	0.56	0.41	0.60	0.50
	(0.06)	(0.04)	(0.05)	(0.06)	(0.10)
$\ln trp_{it}$	0.77	0.34	0.40		
	(0.17)	(0.10)	(0.08)		
$\ln trp_{i,t-1}$	-0.41				
	(0.16)				
$\ln wcap_{it}$	0.25	0.69	0.64	0.67	0.91
	(0.06)	(0.04)	(0.06)	(0.07)	(0.13)
$\ln wcap_{i,t-1}$		-0.32	-0.15	-0.35	-0.43
		(0.06)	(0.07)	(0.08)	(0.17)
sp_{it}		0.39	0.50	0.72	0.36
		(0.09)	(0.07)	(0.10)	(0.07)
$sp_{i,t-1}$	-0.13			-0.32	-0.16
	(0.05)			(0.10)	(0.08)
dv_{it}					
$dv_{i,t-1}$		-0.23			
		(0.08)			
cp_{it}	-0.09	-0.48	-0.13	-0.12	-0.13
	(0.03)	(0.05)	(0.02)	(0.03)	(0.03)
$cp_{i,t-1}$		0.38	0.06	0.07	
		(0.06)	(0.02)	(0.02)	

* Figures in parentheses are bootstrap standard errors (B=100)

Table 9:
LSDVc long-run (semi-) elasticities of specification (5.4) for the five one-digit sectors

	food and agricultural products	textile and leather	chemical and allied products	metal and machinery	electric and electronic manufacturing
trp	0.70	0.80	0.67		
	(0.21)	(0.22)	(0.14)		
wcap	0.46	0.84	0.83	0.82	0.97
	(0.14)	(0.12)	(0.13)	(0.18)	(0.34)
sp	-0.24	0.88	0.83	1.01	0.41
	(0.14)	(0.24)	(0.15)	(0.29)	(0.15)
dv		-0.52			
		(0.26)			
cp	-0.18	-0.23	-0.12	-0.13	-0.26
	(0.07)	(0.17)	(0.03)	(0.07)	(0.08)

* Figures in parentheses are standard errors