# The effects of dynamic feedbacks on LS and MM estimator accuracy in panel data models*

MAURICE J.G. BUN and JAN F. KIVIET
(University of Amsterdam & Tinbergen Institute)

16 October, 2002

*JEL-code:* C13; C23

*Keywords:* asymptotic expansions, bias approximation, dynamic panel data model, feedback mechanisms, Monte Carlo simulation

**Abstract**

The finite sample behaviour is analysed of particular least squares (LS) and method of moments (MM) estimators in panel data models with individual effects and both a lagged dependent variable regressor and another explanatory variable which may be affected by lagged feedbacks from the dependent variable. Asymptotic expansions indicate that the order of magnitude of bias of (generalized) MM estimators tends to increase with the number of moment conditions exploited. For various estimation procedures we examine the analytical effects of feedbacks and other model characteristics such as prominence of individual effects. Simulation results corroborate our theoretical findings and show that in small samples of models with dynamic feedbacks none of the techniques examined dominates. However, a simple bias corrected LS estimator which presupposes strict exogeneity is found to be rather robust, showing often smaller root mean squared errors than GMM.

## 1. Introduction

Economic relationships usually involve dynamic adjustment processes. In regression models it is common practice to model these by including in the specification lagged values of the current regressors, the regressand or both. The inclusion of lagged dependent variables seems to provide an adequate characterization of many economic dynamic adjustment processes, but it induces inference problems such as small sample bias and relegation to possibly poor asymptotic approximations. In dynamic panel data models which allow for unobserved individual effects these problems are aggravated. In such models least-squares (LS), i.e. fixed effects (LSDV) or random effects (GLS), can even be biased in large samples. Although consistent – provided that the disturbances are white-noise and all regressors are predetermined conditional on the individual effects – for a large number of time-series observations $T$, they are inconsistent for $T$ finite and

* Tinbergen Institute & Faculty of Economics and Econometrics, University of Amsterdam, Roetersstraat 11, 1018 WB Amsterdam, The Netherlands (MBUN@FEE.UvA.NL and JFK@FEE.UvA.NL). We thanks Frank Windmeijer who pointed our attention to a flaw in an earlier version of the computer programme that produces our simulation results.

a large number of cross-section observations $N$. Since in micro-economics the typical dimension of a panel data set is a short time span for a large cross-section alternative (generalized) method of moments (MM) estimators have been proposed. These may be inconsistent for $T$ large, but they are consistent for $N$ large (see Anderson and Hsiao, 1982; Arellano and Bond, 1991; Arellano and Bover, 1995; Blundell and Bond, 1998). However, in many actual panel data sets the values of both $N$ and $T$ are only moderately large or even small, for example when in an attempt to mitigate heterogeneity of the slope parameters sub-samples have to be analyzed. In such situations first-order asymptotic approximations seem of little use to indicate which technique is to be preferred.

If in a panel data model for a particular dependent variable one of the explanatory variables is affected by feedbacks from that same dependent variable and this feedback is instantaneous then such an explanatory variable and the regressand are jointly dependent and one should resort to ML (maximum likelihood) or specially designed MM techniques in order to achieve consistency. If this feedback is lagged at least one period and the disturbances are serially uncorrelated then such an explanatory variable, which may show permanent dependence on the unobserved individual effects too, is predetermined. If such feedbacks do not occur then the explanatory variable is strictly exogenous, conditional on the individual effects. In this paper we will focus on the estimation of dynamic panel data models with individual effects and white-noise disturbances, which include a lagged dependent variable and another explanatory variable, and we distinguish the two cases where this latter variable is either strictly exogenous or obeys the weaker property of being predetermined. Hence, we do not consider instantaneous feedbacks, but only lagged feedbacks. The conditional predeterminedness of both regressors implies a range of valid moment conditions and thus defines a range of consistent MM estimators, including the asymptotically most efficient implementation.

Obviously, it would not be wise to use just the standard first-order asymptotic properties (consistency and asymptotic variance) for making a choice between the various LS and MM estimators when the goal is to obtain accurate and efficient inference in samples where both $N$ and $T$ may be small. After all, these asymptotic properties would suggest that there are little worries regarding bias (because all estimators are consistent, either for large $T$ or for large $N$) and among the MM estimators they naturally favor the asymptotically efficient GMM (generalized method of moments) implementation, which uses all available instruments and employs the covariance of the moment conditions as a weighting matrix. It has been observed, though, that in moderately large samples consistent estimators may show substantial bias, and that actual efficiency may deteriorate by using an abundance of moment conditions, see Ziliak (1997) and Koenker and Machado (1999). Therefore, in this study we will apply higher-order asymptotic analysis for various estimators for panel data models with lagged feedbacks, in order to establish the leading term in an expansion of their estimation errors. From that we can obtain the asymptotic order with respect to both $T$ and $N$ of the bias of the various estimators and at the same time we can find the model parameters which seem important for any inaccuracies of the first-order asymptotic approximations. Asymptotically the bias of consistent estimators is invariant with respect to all parameters of the data generating process, but the leading terms in an expansion will contain information on the parameters that determine the bias in finite samples. In addition to these analytical investigations, the actual finite sample bias, standard deviation and mean squared error (MSE) of various LS and MM estimators will be assessed in a series of carefully designed Monte Carlo experiments. The theoretical findings will prove to be helpful in examining any regularities in the determining factors of the actual finite sample characteristics.

In Kiviet (1999) higher-order asymptotic methods have been applied to LSDV and to particular simple MM methods in dynamic panel data models with strictly exogenous or predetermined regressors. These simple MM implementations employ a number of instruments equal to the number of regressors. Hence, they are not asymptotically efficient and moments may not exist. The actual accuracy and relevance of these particular higher-order asymptotic findings have not been examined yet in a Monte Carlo study. Most Monte Carlo studies that have made comparisons between LS and alternative MM estimators in dynamic panel data models examined the case of no or only strictly exogenous regressors. Small $T$ with relatively large $N$ simulation studies are Arellano and Bond (1991), Blundell and Bond (1998), Kiviet (1995) and Alonso-Borrego and Arellano (1999). These show that the bias of LS estimators can be severe, but MM implementations can have a very poor performance too, but less so for larger $N$. A complication of MM estimators is that the instruments may become weak, especially when the coefficient of the lagged dependent variable regressor is large. Judson and Owen (1999) performed simulations with both dimensions of the sample size small or moderate. They find that the bias of LSDV is relatively large compared to simple MM estimators and is more or less equal in magnitude to the bias of GMM. Like Ziliak (1997), they also present some evidence that the bias of MM estimators increases with the number of instruments employed. Generally, it has been established that the variance of LS estimators is relatively small. Therefore, based on a MSE criterion, LSDV estimators which exploit bias correction techniques, as suggested in Kiviet (1995), perform well under a reasonably great number of circumstances. However, clear-cut guidelines for practitioners are not yet available, especially not for the situation where $N$ and $T$ are both small and the regressor next to the lagged dependent variable is predetermined and not strictly exogenous. Some Monte Carlo results for models with instantaneous feedbacks are presented in Blundell, Bond and Windmeijer (2000), but simulation evidence for models with various forms of lagged feedbacks as considered here was not yet available.

Hence, we shall examine the accuracy of inference obtained by various MM implementations, especially the GMM procedures which exploit a number of moment conditions of order $O(T^2)$, and make comparisons with results for LSDV and GLS. We also consider MM techniques which exploit fewer instruments, viz. $O(T)$, and find important qualitative and numerical differences between estimators regarding their order of magnitude of small sample bias and efficiency. One general finding is that the bias of all estimators can be substantial and is affected by lagged feedbacks via the explanatory variable in a similar manner as by the inclusion of the lagged dependent variable. Therefore, approximation of this bias in practice would require adequate specification and consistent estimation of all feedback relationships. Because this does not seem operational in general, we explore an easily applicable (but rather naive) approach and check whether the simple corrected LSDV estimator, in which we (possibly incorrectly) adopt strict exogeneity of the regressor, works well and we find that it often beats most MM estimators.

The structure of this paper is as follows. After the introduction of the model and its particular stochastic structure in Section 2 and the various estimators in Section 3, we obtain in Section 4 the leading terms of the estimation errors of the various estimators. There we establish marked differences between all the estimators examined regarding the order of magnitude of their bias in terms of $N$ and $T$. At the same time the effects on estimator location resulting from a lagged feedback instead of strict exogeneity of the explanatory variable are made explicit and so is the finite sample dependence of some estimators regarding the prominence of the unobserved individual effects. These are shown to limit the options for bias correction procedures, which we discus in Section 5. In Section 6 an appropriate simulation design is constructed and experiments are performed to examine whether the analytical higher-order asymptotic results are corroborated by the actual performance of the estimators in small samples. Section 7 concludes.

## 2. Stochastic structure of the model

We will consider estimation methods for the standard first-order dynamic panel data model with random individual effects and only one further explanatory variable, i.e.

$$y_{it} = \gamma y_{i,t-1} + \beta x_{it} + u_{it}, \qquad i = 1, ..., N; \; t = 1, ..., T. \tag{2.1}$$

The disturbance term

$$u_{it} = \eta_i + \varepsilon_{it}, \tag{2.2}$$

contains two error components, viz. an unobserved individual specific effect $\eta_i$ and a general disturbance term $\varepsilon_{it}$. In this model the dependent variable $y_{it}$ depends on its one period lagged value and on a time variant explanatory variable $x_{it}$. We assume that the regressor $x_{it}$ may be correlated with $\eta_i$ and is predetermined with respect to $\varepsilon_{it}$, i.e.

$$\left.\begin{array}{ll} \mathsf{E}(x_{is}\varepsilon_{it}) = 0, & s \leq t \\ \mathsf{E}(x_{is}\varepsilon_{it}) \neq 0, & s > t \end{array}\right\} \qquad i = 1, ..., N. \tag{2.3}$$

Below we will formalize the correlation of $x_{it}$ with $\eta_i$ and the lagged feedback mechanism of past disturbances on $x_{it}$. Although the LS and MM estimators to be examined do not necessitate to specify a distribution function for the error components, our higher-order asymptotic derivations shall require some higher-order moments. For the sake of simplicity we will be more specific on this than strictly required and assume

$$\left.\begin{array}{l} \eta_i \sim \text{i.i.d.N}(0, \sigma_\eta^2) \\ \varepsilon_{it} \sim \text{i.i.d.N}(0, \sigma_\varepsilon^2) \end{array}\right\} \qquad i = 1, ..., N; \; t = 0, ..., T. \tag{2.4}$$

We define $\varepsilon_{i0}$ because it enables to specify the random characteristics of the start-up values $y_{i0}$ and lagged feedbacks in $x_{i1}$, as we shall see. As usual we assume that the two error components are uncorrelated, i.e.

$$\mathsf{E}(\eta_i \varepsilon_{jt}) = 0, \qquad \forall i, j, t, \tag{2.5}$$

and that

$$\mathsf{E}(y_{i0}\varepsilon_{jt}) = 0, \qquad \forall i, j, t > 0, \tag{2.6}$$

i.e. all $N$ initial observations $y_{i0}$ are uncorrelated with all disturbances for $t > 0$. Furthermore, we suppose that the model in (2.1) is dynamically stable, i.e. $|\gamma| < 1$.

Following Kiviet (1999) we decompose $y$ and $x$ into a zero-mean relevant random component, denoted by a tilde, and irrelevant random plus deterministic components, denoted by a bar. The relevant random components are those which are related to the individual effects $\eta_i$ and the disturbance terms $\varepsilon_{it}$. Regarding $x$ we start by using the same simple setup as in Kiviet (1999), i.e. $x_{it}$ is decomposed as

$$\left.\begin{aligned} x_{it} &= \bar{x}_{it} + \tilde{x}_{it} \\ \tilde{x}_{it} &= \phi\varepsilon_{i,t-1} + \pi\eta_i \end{aligned}\right\} \qquad i = 1, ..., N;\ t = 1, ..., T, \tag{2.7}$$

where $\mathsf{E}(\bar{x}_{it}\eta_j) = 0$ and $\mathsf{E}(\bar{x}_{it}\varepsilon_{js}) = 0$ for $\forall i, j, t, s$. The parameter $\pi$ allows for correlation between observed and unobserved heterogeneity and the parameter $\phi$ determines the feedback of the lagged disturbance into the explanatory variable $x_{it}$.

For the relevant random component $\tilde{y}_{it}$ of $y_{it}$ and its complement $\bar{y}_{it}$ we have

$$\left.\begin{aligned} \tilde{y}_{it} &= \gamma\tilde{y}_{i,t-1} + \beta\tilde{x}_{it} + \eta_i + \varepsilon_{it} \\ \bar{y}_{it} &= \gamma\bar{y}_{i,t-1} + \beta\bar{x}_{it} \end{aligned}\right\} \qquad i = 1, ..., N;\ t = 1, ..., T. \tag{2.8}$$

In the analysis to follow we shall condition on $\bar{x}_{it}$, $\bar{y}_{it}$, $t = 1, ..., T$ and on $\bar{y}_{i0} = y_{i0} - \tilde{y}_{i0}$, for all $i$. To be able to decompose the relevant random components of $\tilde{y}_{it}$ into the two error components $\eta_i$ and $\varepsilon_{it}$ we adopt

$$\mathsf{E}(\tilde{y}_{it} \mid \eta_i) = \alpha\eta_i, \qquad i = 1, ..., N;\ t = 0, ..., T \tag{2.9}$$

where $\alpha = \frac{1+\beta\pi}{1-\gamma}$. Hence, we assume that the full long-run impact of the individual effect $\eta_i$ on $y_{it}$ is already present in $y_{i0}$. For the initial values we further assume

$$\tilde{y}_{i0} = \alpha\eta_i + \omega\varepsilon_{i0}, \qquad \forall i \tag{2.10}$$

where $\omega$ is either 0 or 1. In case of lagged feedback $x_{i1}$ depends on $\varepsilon_{i0}$ and then $\tilde{y}_{i0}$ should also contain $\varepsilon_{i0}$, because it is in this case a relevant random component of $y_{i0}$. Hence, when $\phi \neq 0$ we take $\omega = 1$ giving $\tilde{y}_{i0} = \alpha\eta_i + \varepsilon_{i0}$. However, we choose to take $\omega = 0$ in case $\phi = 0$, because when $x_{it}$ is strictly exogenous the normal procedure is to condition on $x_{it}$ and on $\bar{y}_{i0} = y_{i0} - \alpha\eta_i$. Hence, in that case $\tilde{y}_{i0}$ should not contain $\varepsilon_{i0}$ (because it is an irrelevant random component now), thus $\tilde{y}_{i0} = \alpha\eta_i$ when $\phi = 0$.

Stacking the observations over time we get $(i = 1, ..., N)$

$$y_i = \gamma y_{i(-1)} + \beta x_i + \eta_i\iota_T + \varepsilon_i, \tag{2.11}$$

$$x_i = \bar{x}_i + \phi\varepsilon_{i(-1)} + \pi\eta_i\iota_T, \tag{2.12}$$

where $y_{i(-1)} = (y_{i0}, ..., y_{i,T-1})'$, $\varepsilon_{i(-1)} = (\varepsilon_{i0}, ..., \varepsilon_{i,T-1})'$ and $\iota_T = (1, ..., 1)'$ a $T \times 1$ vector of ones. From the above it follows that

$$\begin{aligned} \tilde{y}_i &= \gamma\tilde{y}_{i(-1)} + \beta\phi\varepsilon_{i(-1)} + (\beta\pi + 1)\eta_i\iota_T + \varepsilon_i \\ &= \gamma(L_T\tilde{y}_i + \tilde{y}_{i0}e_{T,1}) + \beta\phi(L_T\varepsilon_i + \varepsilon_{i0}e_{T,1}) + (\beta\pi + 1)\eta_i\iota_T + \varepsilon_i, \end{aligned}$$

where we introduced the $T \times T$ matrix $L_T$ with ones on the first lower subdiagonal and zeros elsewhere and the $q \times 1$ unit vector $e_{q,p}$ with $p^{th}$ element equal to one. Defining

$$\Gamma_T = (I_T - \gamma L_T)^{-1} \tag{2.13}$$

and using (2.10), the relevant random part of $y_i$ can now be written explicitly in terms of the error components as

$$\tilde{y}_i = \alpha\eta_i\iota_T + \Gamma_T(I_T + \beta\phi L_T)\varepsilon_i + (\omega\gamma + \phi\beta)\Gamma_T e_{T,1}\varepsilon_{i0}. \tag{2.14}$$

Stacking the $T$ observations per individual over all $N$ individuals yields

$$y = W\delta + u, \tag{2.15}$$

where $\delta = (\gamma, \beta)'$, $y$ and $u$ are $NT \times 1$, $W = (y_{(-1)}, x)$ is $NT \times 2$, $u = S\eta + \varepsilon$, with $S = I_N \otimes \iota_T$ an $NT \times N$ matrix and $\eta = (\eta_1, ..., \eta_N)'$. Using (2.7) and (2.14) the relevant random parts of $y$ and $x$ can be written as

$$\tilde{y} = \alpha S\eta + \Gamma(I_{NT} + \beta\phi L)\varepsilon + (\omega\gamma + \phi\beta)\Gamma(I_N \otimes e_{T,1})\varepsilon_0, \tag{2.16}$$

and

$$\tilde{x} = \pi S\eta + \phi[L\varepsilon + (I_N \otimes e_{T,1})\varepsilon_0], \tag{2.17}$$

where $\Gamma = I_N \otimes \Gamma_T$, $L = I_N \otimes L_T$ and $\varepsilon_0 = (\varepsilon_{10}, ..., \varepsilon_{N0})'$.

The above expressions are required for obtaining analytical results on the properties of the various estimators of $\delta$ introduced in the next section. To express some of these estimators particular transformations of the data have to be considered, for which we will use the $(T-1) \times T$ transformation matrices

$$J_T = \begin{pmatrix} 0 & 1 & 0 & \cdot & \cdot & 0 \\ 0 & 0 & 1 & \cdot & & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & & & \cdot & 1 & 0 \\ 0 & \cdot & \cdot & \cdot & 0 & 1 \end{pmatrix}, \quad K_T = \begin{pmatrix} 1 & 0 & \cdot & \cdot & \cdot & 0 \\ 0 & 1 & 0 & \cdot & & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & & \cdot & 1 & 0 & \cdot \\ 0 & \cdot & \cdot & 0 & 1 & 0 \end{pmatrix} \tag{2.18}$$

and also $D_T = J_T - K_T$. Note that $D_T$ transforms a $T$ element vector for individual $i$ into $T-1$ first differences, because $J_T$ skips the first observation and $K_T$ skips the final observation. We define also $D = I_N \otimes D_T$, $J = I_N \otimes J_T$ and $K = I_N \otimes K_T$.

# 3. Estimators for dynamic panel data models

## 3.1. Least Squares

Treating the random individual specific effects as fixed, estimation of $\delta$ and $\eta$ in (2.15) by OLS yields estimates which are called Least Squares Dummy Variables (LSDV), fixed effect or within group estimates. For $\delta$ this estimator can be expressed as

$$\hat{\delta}_{LSDV} = (W'AW)^{-1}W'Ay, \tag{3.1}$$

where the $NT \times NT$ matrix $A = I_N \otimes A_T$ with $A_T = I_T - \frac{1}{T}\iota_T\iota_T'$ is the within transformation which wipes out the individual effects. Estimator (3.1) can be written also as

$$\hat{\delta}_{LSDV} = (W^{*\prime}W^*)^{-1}W^{*\prime}y^*, \tag{3.2}$$

where $y^* = Py$, $W^* = PW$ and $P = I_N \otimes P_T$ is the forward orthogonal deviations operator, see Arellano and Bover (1995). This transformation will prove to be useful especially when constructing and analyzing MM estimators. The $(T-1) \times T$ upper-triangular matrix $P_T$ transforms as

$$y^*_{it} = c_t[y_{it} - (T-t)^{-1}(y_{i,t+1} + ... + y_{iT})], \tag{3.3}$$

with $c_t^2 = (T-t)/(T-t+1)$, whereas $P_T P_T' = I_{T-1}$, $P_T' P_T = A_T$ and $P_T = (D_T D_T')^{-1/2} D_T$, where $D_T$ is the first difference operator as defined above. Note that the columns of $D_T'$ span the orthogonal complement of $\iota_T$. Hence, projection on the orthogonal complement of $\iota_T$ conforms to projection on $D_T'$ or

$$A_T = I_T - \iota_T(\iota_T'\iota_T)^{-1}\iota_T' = D_T'(D_T D_T')^{-1} D_T. \tag{3.4}$$

By choosing an upper-triangular matrix for $(D_T D_T')^{-1/2}$, the transformation (3.3) follows. When applied to i.i.d. disturbances, the transformation $P_T$ preserves the independence, hence it is referred to as orthogonal deviations.

Treating the individual effects as random, the covariance matrix of the disturbance term $u = S\eta + \varepsilon$ is

$$V = \sigma_\eta^2 SS' + \sigma_\varepsilon^2 I_{NT} \tag{3.5}$$

and the Generalized Least Squares (GLS) estimator of $\delta$ is

$$\hat{\delta}_{GLS} = (W'V^{-1}W)^{-1}W'V^{-1}y, \tag{3.6}$$

where

$$V^{-1} = I_N \otimes (I_T - \theta \frac{1}{T}\iota_T\iota_T'), \qquad \text{with } \theta = 1 - \left(1 + T\frac{\sigma_\eta^2}{\sigma_\varepsilon^2}\right)^{-1}. \tag{3.7}$$

Note that $0 \leq \theta \leq 1$, and $\theta \to 1$ as $T \to \infty$ for $\sigma_\eta^2 > 0$, which shows the equivalence of the GLS and LSDV estimators for $T$ large. However, for finite $T$ both estimators differ and $V^{-1}$ will only partially wipe out the individual effects (which explains that we will establish that the finite sample bias of GLS, unlike that of LSDV, is not invariant with respect to $\sigma_\eta^2/\sigma_\varepsilon^2$).

### 3.2. Method of Moments

The assumptions made on the stochastic structure of the model imply for each individual $i$ a set of linear and non-linear moment conditions. In this study we will focus on MM implementations using linear moment conditions only, see Arellano and Bond (1991) and Blundell and Bond (1998), and we abstain from orthogonality conditions arising from the i.i.d. assumption of the disturbances $\varepsilon_{it}$ as proposed by Ahn and Schmidt (1995).

Different sets of moment conditions are available depending on whether or how the individual effects are removed from the model. Arellano and Bond (1991) suggest to take first differences, i.e.

$$Dy = DW\delta + D\varepsilon. \tag{3.8}$$

Since $\mathsf{E}[(Dy_{(-1)})'D\varepsilon] = -\mathsf{E}(y_{(-1)}'\varepsilon_{(-1)}) = -\mathsf{E}(\varepsilon_{(-1)}'\varepsilon_{(-1)}) = -\sigma_\varepsilon^2 N(T-1)$, the correlation of one of the regressors with the errors of (3.8) is such that here OLS is inconsistent, irrespective of how the sample size is extended. Valid moment conditions for individual $i$ in the differenced model (3.8) are

$$\left.\begin{array}{lll} \mathsf{E}(y_{i,t-s}\Delta\varepsilon_{it}) = 0, & t = 2,...,T; & s = 2,...,t, \\ \mathsf{E}(x_{i,t-s}\Delta\varepsilon_{it}) = 0, & t = 2,...,T; & s = 1,...,t-1. \end{array}\right\} \tag{3.9}$$

Of course, even more valid instruments than these $T(T-1)$ are available when $x$ is strictly exogenous, viz. $\mathsf{E}(x_{is}\Delta\varepsilon_{it}) = 0$ for $t = 2,...,T$; $s = 1,...,T$, but we will not use these here.

### 3.2.1. Removing the effects from the model

In line with Arellano and Honoré (2001, p.3255) we shall consider any $(T-1)\times T$ upper triangular matrix transformation matrix $B_T$ with rank $(T-1)$ and $B_T\iota_T = 0$, such as for example $D_T$ or $P_T$ as defined earlier, but will not necessarily exploit all linear moment restrictions. Hence, we consider the model

$$By = BW\delta + B\varepsilon, \tag{3.10}$$

where transformation $B = I_N \otimes B_T$ dissolves the individual effects, and consider GMM estimation with (a subset of) $m$ moment conditions for each individual $i$, which can be expressed as

$$\mathsf{E}(Z_{li}' B_T \varepsilon_i) = 0. \tag{3.11}$$

Here $Z_{li}$ is a $(T-1)\times m$ block-diagonal matrix with variables in levels. We will consider two particular instrument matrices in this study, viz.

$$Z_{li}^{(1)} = \begin{pmatrix} y_{i0} & x_{i1} & 0 & 0 & 0' & \cdots & 0 & 0 \\ 0 & 0 & y_{i1} & x_{i2} & 0' & & & 0 \\ \vdots & \vdots & & & \cdots & & & \vdots \\ 0 & 0 & 0 & 0 & & \cdots & 0 & 0 \\ 0 & 0 & 0 & 0 & \cdots & 0' & y_{i,T-2} & x_{i,T-1} \end{pmatrix}$$

and

$$Z_{li}^{(2)} = \begin{pmatrix} y_{i0} & x_{i1} & 0 & 0 & 0 & 0 & 0' & \cdots & 0 & \cdots & 0 & 0 & \cdots & 0 \\ 0 & 0 & y_{i0} & y_{i1} & x_{i1} & x_{i2} & 0' & \cdots & 0 & & & & & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \cdots & & \vdots & & & & & \vdots \\ \vdots & & & & & \vdots & & \cdots & 0 & \cdots & 0 & 0 & \cdots & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \cdots & \cdots & y_{i0} & \cdots & y_{i,T-2} & x_{i1} & \cdots & x_{i,T-1} \end{pmatrix}.$$

The matrix $Z_{li}^{(2)}$ includes all $m = T(T-1) = O(T^2)$ instruments given in (3.9), while $Z_{li}^{(1)}$ includes only a subset of $m = 2*(T-1) = O(T)$ of these.

Stacking over individuals one gets for any instrument matrix $Z_l$

$$\mathsf{E}(Z_l' B\varepsilon) = 0, \tag{3.12}$$

where $Z_l = (Z_{l1}', ..., Z_{lN}')'$. Provided $B'Z_l$ has full column rank, i.e. $m \leq NT$, the one-step GMM estimator based on $\mathsf{E}(Z_l'B\varepsilon) = 0$ is

$$\hat{\delta}_{GMM} = [W'B'Z_l(Z_l'BB'Z_l)^{-1}Z_l'BW]^{-1}W'B'Z_l(Z_l'BB'Z_l)^{-1}Z_l'By, \tag{3.13}$$

since $(Z_l'BB'Z_l)^{-1}$ exists and is a consistent estimate of the inverse of $\mathsf{E}(Z_l'B\varepsilon\varepsilon'B'Z_l)$ up to a scalar under the i.i.d. assumption for $\varepsilon_{it}$. Therefore estimator (3.13) is asymptotically efficient in the class of MM estimators when all moment conditions (3.9) are exploited, i.e. $Z_l = Z_l^{(2)}$. Then for $B = D$ the estimator (3.13) specializes to the one-step estimator of Arellano and Bond (1991) and for $B = P$ it specializes to the GMM estimator analyzed in Alvarez and Arellano (1998), which is in fact equivalent (proof in Appendix A). Since the latter form has proved to be more suitable for further analytical examination we will focus in this study on transformation $P$ too, investigating it also for $Z_l = Z_l^{(1)}$ in order to verify the effects on small sample properties of a reduction in the number of instruments exploited. We will indicate these estimators by $\hat{\delta}^{(1)}_{GMMpl}$ and $\hat{\delta}^{(2)}_{GMMpl}$ respectively, clarifying which matrix $Z_l$ has been used and that the model underwent the $P$ transformation. The requirement $m \leq NT$ has no noteworthy implications for $\hat{\delta}^{(1)}_{GMMpl}$, whereas it imposes $N \geq T-1$ for $\hat{\delta}^{(2)}_{GMMpl}$.

For $B = P$ and writing $y^* = Py$ and $W^* = PW$ estimator (3.13) can be written as

$$\hat{\delta}_{GMMpl} = [W^{*\prime} Z_l (Z_l' Z_l)^{-1} Z_l' W^*]^{-1} W^{*\prime} Z_l (Z_l' Z_l)^{-1} Z_l' y^*. \tag{3.14}$$

Now let $Q$ be an $N(T-1) \times N(T-1)$ permutation matrix, which changes the order of the rows of $Z_l$, $W^*$ and $y^*$ such that no longer $T-1$ rows for the separate individuals are put together in $N$ sub-matrices, but $T-1$ sub-matrices of $N$ rows (the initial observation of all individuals on top, etc.). Then $QZ_l = (Z_{l1}', ..., Z_{l,T-1}')'$ and this is now a block-diagonal matrix, because $Z_{li}^{(1)}$ and $Z_{li}^{(2)}$ are. Of course, $Q$ is orthonormal, i.e. $Q'Q = I_{N(T-1)}$ and $Q^{-1} = Q'$, hence

$$(Z_l' Z_l)^{-1} = (Z_l' Q' Q Z_l)^{-1} = diag[(Z_{l1}' Z_{l1})^{-1}, ..., (Z_{l,T-1}' Z_{l,T-1})^{-1}],$$

so that one can write, see also Arellano and Bover (1995),

$$\hat{\delta}_{GMMpl} = \left[\sum_{t=1}^{T-1} W_t^{*\prime} Z_{lt} (Z_{lt}' Z_{lt})^{-1} Z_{lt}' W_t^*\right]^{-1} \sum_{t=1}^{T-1} W_t^{*\prime} Z_{lt} (Z_{lt}' Z_{lt})^{-1} Z_{lt}' y_t^*, \tag{3.15}$$

where $y_t^* = (y_{1t}^*, ..., y_{Nt}^*)'$ is a $N \times 1$ vector, $W_t^* = (w_{1t}^*, ..., w_{Nt}^*)'$ a $N \times 2$ matrix and $Z_{lt}$ a $N \times m_{lt}^{(r)}$ matrix, $r = \{1, 2\}$. In case of using all available instruments, i.e. $Z_l = Z_l^{(2)}$, we have $m_{lt}^{(2)} = 2t$, while in case of $Z_l = Z_l^{(1)}$ we have $m_{lt}^{(1)} = 2$.

### 3.2.2. Removing the effects from the instruments

For the equation in levels (2.15) first differences (and their lags) of the explanatory variables are valid instruments, see Arellano and Bover (1995), Kiviet (1995) and Blundell and Bond (1998). The additional moment conditions can be combined with (3.9) into a system GMM estimator. In this study, however, we analyze the simpler GMM estimator based on instruments in first differences only, i.e. exploiting

$$\left.\begin{array}{lll} \mathsf{E}(u_{it}\Delta y_{i,t-s}) = 0, & t = 2, ..., T; & s = 1, ..., t-1, \\ \mathsf{E}(u_{it}\Delta x_{i,t-s}) = 0, & t = 2, ..., T; & s = 0, ..., t-2. \end{array}\right\} \tag{3.16}$$

The set of moment conditions for individual $i$ can be expressed as (Blundell, Bond and Windmeijer, 2000)

$$\mathsf{E}(Z_{di}' J_T u_i) = 0, \tag{3.17}$$

where the set of $m$ linear moment conditions is embodied in the $(T-1) \times m$ matrix $Z_{di}$. We will consider the following particular instrument matrices

$$Z_{di}^{(1)} = \begin{pmatrix} \Delta y_{i1} & \Delta x_{i2} & 0' & \cdots & 0 & 0 \\ 0 & 0 & \cdots & & & 0 \\ \vdots & \vdots & & & & \vdots \\ 0 & 0 & & \cdots & 0 & 0 \\ 0 & 0 & \cdots & 0' & \Delta y_{i,T-1} & \Delta x_{i,T} \end{pmatrix}$$

and

$$Z_{di}^{(2)} = \begin{pmatrix} \Delta y_{i1} & \Delta x_{i2} & 0' & \cdots & 0 & \cdots & 0 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & \cdots & 0 & & & & & 0 \\ 0 & 0 & \cdots & & \vdots & & & & & \vdots \\ \vdots & & & \cdots & 0 & \cdots & 0 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & \cdots & \Delta y_{i1} & \cdots & \Delta y_{i,T-1} & \Delta x_{i2} & \cdots & \Delta x_{i,T} \end{pmatrix}.$$

The matrix $Z_{di}^{(2)}$ includes all $m^{(2)} = T(T-1) = O(T^2)$ instruments which are available when $x$ is predetermined, while $Z_{di}^{(1)}$ contains only a subset of $2(T-1) = O(T)$ of these. Stacking over individuals one gets

$$\mathsf{E}(Z_d' J u) = 0, \tag{3.18}$$

where $Z_d = (Z_{d1}', ..., Z_{dN}')'$ is either $Z_d^{(1)}$ or $Z_d^{(2)}$. Note that in this case one-step estimators based on $\mathsf{E}(Z_d' J u) = 0$ will not be asymptotically equivalent with two-step estimators unless $\sigma_\eta^2 = 0$. Here, we will analyze a particular one-step GMM estimator, which has been analyzed by Blundell, Bond and Windmeijer (2000) and has been exploited also in the construction of a system GMM estimator (Blundell and Bond, 1998). This GMM estimator uses as a weighting matrix $(Z_d' Z_d)^{-1}$, provided this inverse exists i.e. $m \leq N(T-1)$. It can then be written as

$$\hat{\delta}_{GMMld} = [W'J'Z_d(Z_d'Z_d)^{-1}Z_d'JW]^{-1}W'J'Z_d(Z_d'Z_d)^{-1}Z_d'Jy, \tag{3.19}$$

clarifying that the equation in levels has been instrumented by differenced variables. Employing the permutation $Q$ again it can also be written as

$$\hat{\delta}_{GMMld} = \left[\sum_{t=2}^{T} W_t' Z_{dt}(Z_{dt}'Z_{dt})^{-1}Z_{dt}'W_t\right]^{-1} \sum_{t=2}^{T} W_t' Z_{dt}(Z_{dt}'Z_{dt})^{-1}Z_{dt}'y_t, \tag{3.20}$$

where $y_t = (y_{1t}, ..., y_{Nt})'$ is a $N \times 1$ vector, $W_t = (w_{1t}, ..., w_{Nt})'$ a $N \times 2$ matrix and $Z_{dt}$ a $N \times m_{dt}^{(r)}$ matrix, $r = \{1, 2\}$. When $Z_d^{(2)}$ is used $N \geq T$ is required, $m_{dt}^{(2)} = 2t$ and the estimator is denoted by $\hat{\delta}_{GMMld}^{(2)}$. When $Z_d^{(1)}$ is used $m_{dt}^{(1)} = 2$ giving $\hat{\delta}_{GMMld}^{(1)}$.

## 4. Finite sample bias

In this section we will focus on the location of LSDV, GLS and GMM estimators of $\delta$. More in particular, we will approximate the finite sample bias of the various estimators using asymptotic expansion techniques.

The estimation error of the LSDV estimator (3.1), i.e.

$$\hat{\delta}_{LSDV} - \delta = (W'AW)^{-1} W'A\varepsilon, \tag{4.1}$$

depends on $\varepsilon$ in a complicated non-linear way. However, we can expand as follows. Consider

$$\begin{aligned} W'AW &= \Phi^{-1} + W'AW - \Phi^{-1} \\ &= [I + (W'AW - \Phi^{-1})\Phi]\Phi^{-1}, \end{aligned} \tag{4.2}$$

where $\Phi^{-1} = \mathsf{E}(W'AW)$ is of order $O(n)$ with $n = NT$, see Kiviet (1995, 1999). Also $W'AW = O_p(n)$ and $W'AW - \Phi^{-1} = O_p(n^{1/2})$, hence

$$\begin{aligned} (W'AW)^{-1} &= \Phi[I + (W'AW - \Phi^{-1})\Phi]^{-1} \\ &= \Phi[I - (W'AW - \Phi^{-1})\Phi + ...] \\ &= \Phi + O_p(n^{-3/2}), \end{aligned} \tag{4.3}$$

and because $W'A\varepsilon = O_p(n^{1/2})$,

$$\mathsf{E}(\hat{\delta}_{LSDV} - \delta) = \Phi\mathsf{E}(W'A\varepsilon) + O(n^{-1}). \tag{4.4}$$

In Appendix B it is shown that the expected estimation errors of the GLS estimator (3.6) and the GMM estimators (3.14) and (3.19) can be expressed as

$$\mathsf{E}(\hat{\delta}_{GLS} - \delta) = \Phi_{GLS}\mathsf{E}(W'V^{-1}u) + O(n^{-1}) \tag{4.5}$$

$$\mathsf{E}(\hat{\delta}_{GMMpl} - \delta) = \Upsilon\mathsf{E}(W'P'M_{Z_l}P\varepsilon) + O(n^{-1}) \tag{4.6}$$

$$\mathsf{E}(\hat{\delta}_{GMMld} - \delta) = \Theta\mathsf{E}(W'J'M_{Z_d}Ju) + O(n^{-1}), \tag{4.7}$$

with $M_Z = Z(Z'Z)^{-1}Z'$ and where $\Phi_{GLS}^{-1} = \mathsf{E}(W'V^{-1}W)$, $\Upsilon^{-1} = \mathsf{E}(W'P'M_{Z_l}PW)$ and $\Theta^{-1} = \mathsf{E}(W'J'M_{Z_d}JW)$ are all of order $O(n)$.

We will analyze the leading bias terms in (4.4) through (4.7) in more detail and establish the order of magnitude of the leading term for the various estimators (proofs in Appendix C). As $\Phi$, $\Phi_{GLS}$, $\Upsilon$ and $\Theta$ are all $O(n^{-1})$, differences in the order of magnitude of the bias of the various estimators will depend on differences in the orders of $\mathsf{E}(W'A\varepsilon)$, $\mathsf{E}(W'V^{-1}u)$, $\mathsf{E}(W'P'M_{Z_l}P\varepsilon)$ and $\mathsf{E}(W'J'M_{Z_d}Ju)$, which vary as we shall see, although $W'A\varepsilon$, $W'V^{-1}u$, $W'P'M_{Z_l}P\varepsilon$ and $W'J'M_{Z_d}Ju$ are all $O_p(n^{1/2})$.

For the LSDV estimator in the present model Kiviet (1999) obtained

$$\mathsf{E}(\hat{\delta}_{LSDV} - \delta) = B_{LSDV}(T^{-1}) + O(n^{-1}), \tag{4.8}$$

with

$$B_{LSDV}(T^{-1}) = \sigma_\varepsilon^2[\mathrm{tr}(\Pi) + \beta\phi\,\mathrm{tr}(\Pi L)]\Phi e_{2,1} + \sigma_\varepsilon^2\phi\,\mathrm{tr}(AL)\Phi e_{2,2} = O(T^{-1}), \tag{4.9}$$

where $\Pi = AL\Gamma$. This result highlights that the leading term of the bias of LSDV is affected by the feedback parameter $\phi$ (note that $\Phi$ is affected by $\phi$ too).

For the GLS estimator we find

$$\mathsf{E}(\hat{\delta}_{GLS} - \delta) = B_{GLS}(T^{-1}) + O(n^{-1}), \tag{4.10}$$

with

$$\begin{aligned} B_{GLS}(T^{-1}) = {} & \sigma_\varepsilon^2\left[\mathrm{tr}(V^{-1}L\Gamma) + \beta\phi\,\mathrm{tr}(V^{-1}L\Gamma L) + \alpha\frac{NT\frac{\sigma_\eta^2}{\sigma_\varepsilon^2}}{1 + T\frac{\sigma_\eta^2}{\sigma_\varepsilon^2}}\right]\Phi_{GLS}e_{2,1} \\ & +\sigma_\varepsilon^2\left[\phi\,\mathrm{tr}(V^{-1}L) + \pi\frac{NT\frac{\sigma_\eta^2}{\sigma_\varepsilon^2}}{1 + T\frac{\sigma_\eta^2}{\sigma_\varepsilon^2}}\right]\Phi_{GLS}e_{2,2} = O(T^{-1}). \end{aligned} \tag{4.11}$$

Hence, the bias of GLS depends on $\gamma$, $\beta$, $\phi$ and $\sigma_\varepsilon^2$, as is the case for LSDV, but the leading bias term of GLS is affected too by the relative magnitude of the error components $\sigma_\eta^2/\sigma_\varepsilon^2$

and by the dependence of the regressor $x$ on the individual effects as parametrized by $\pi$. So, the LSDV estimator seems more robust than GLS.

Next, we consider the GMM estimators using instruments in levels, i.e. $\hat{\delta}^{(2)}_{GMMpl}$ and $\hat{\delta}^{(1)}_{GMMpl}$, and find

$$\mathsf{E}(\hat{\delta}^{(2)}_{GMMpl} - \delta) = B^{(2)}_{GMMpl}(N^{-1}) + O(n^{-1}), \tag{4.12}$$

with

$$\begin{aligned} B^{(2)}_{GMMpl}(N^{-1}) &= 2\sigma^2_\varepsilon \sum_{s=2}^{T} \mathrm{tr}[H_s A_s H_s' L_T \Gamma_T (I_T + \beta\phi L_T)]\Upsilon e_{2,1} \\ &\quad + 2\sigma^2_\varepsilon \phi \sum_{s=2}^{T} \mathrm{tr}(H_s A_s H_s' L_T)\Upsilon e_{2,2} = O(N^{-1}), \end{aligned} \tag{4.13}$$

where $H_s = (0 : I_s)'$. Regarding the bias of the GMM$pl$ estimator, which uses only a number of moment conditions of order $T$ (instead of $T^2$), we find

$$\mathsf{E}(\hat{\delta}^{(1)}_{GMMpl} - \delta) = B^{(1)}_{GMMpl}(n^{-1}) + O(n^{-1}), \tag{4.14}$$

with

$$B^{(1)}_{GMMpl}(n^{-1}) = 2\sigma^2_\varepsilon \,\mathrm{tr}[\Pi_T(I_T + \beta\phi L_T)]\Upsilon e_{2,1} + 2\sigma^2_\varepsilon \phi \,\mathrm{tr}(A_T L_T)\Upsilon e_{2,2} = O(n^{-1}). \tag{4.15}$$

Hence, reducing the number of instruments by a factor $T$ has also reduced the leading term of the bias by a factor $T$.

Finally, we consider the GMM estimators using instruments in first differences for the levels equation, i.e. $\hat{\delta}^{(2)}_{GMMld}$ and $\hat{\delta}^{(1)}_{GMMld}$, and find

$$\mathsf{E}(\hat{\delta}^{(2)}_{GMMld} - \delta) = B^{(2)}_{GMMld}(TN^{-1}) + O(n^{-1}), \tag{4.16}$$

where

$$B^{(2)}_{GMMld}(TN^{-1}) = \sigma^2_\eta (T^2 + T - 2)\Theta(\alpha e_{2,1} + \pi e_{2,2}) = O(TN^{-1}). \tag{4.17}$$

Furthermore, we find

$$\mathsf{E}(\hat{\delta}^{(1)}_{GMMld} - \delta) = B^{(1)}_{GMMld}(N^{-1}) + O(n^{-1}), \tag{4.18}$$

with

$$B^{(1)}_{GMMld}(N^{-1}) = \sigma^2_\eta (T - 1)\Theta(\alpha e_{2,1} + \pi e_{2,2}) = O(N^{-1}), \tag{4.19}$$

hence again the bias has reduced by a factor $T$.

The results on finite sample bias in models with predetermined regressors obtained here have been summarized in Table 1. The columns of this table indicate the model formulation and other characteristics. The leading bias terms of the LSDV and GLS estimators is $O(T^{-1})$. Note that applying GLS to the levels equation is equivalent with GLS applied to the first-differenced model, and LSDV conforms to applying LS after the orthogonal deviations transformation of the model. Although both LS estimators have a bias with leading term of similar order regarding the sample size, we have also assessed that these leading terms have different invariance properties, the GLS estimator being less robust than LSDV as its estimation error depends on both $\sigma_\eta^2/\sigma_\varepsilon^2$ and $\pi$. The GMM$pl^{(2)}$ estimator, which uses all instruments, is equivalent for the orthogonal deviations and first difference equations and has an $O(N^{-1})$ leading bias term, as was already found by Alvarez and Arellano (1998) for the panel AR(1) model. However, when using only a number of instruments of order $T$ the leading bias term of the GMM$pl^{(1)}$ estimator is $O(T^{-1}N^{-1})$, reflecting the noteworthy fact that using fewer moment conditions reduces the order of magnitude of the finite sample bias too. We also find that the GMM$ld^{(2)}$ estimator has a leading bias term of order $O(TN^{-1})$, which reduces to $O(N^{-1})$ for GMM$ld^{(1)}$, but in addition the bias of this estimator of the untransformed model depends heavily on $\sigma_\eta^2$ and $\pi$. Note that for similar number of instruments the bias of GMM$pl$ is a factor $T$ smaller[1] than for GMM$ld$.

These results indicate that, as far as bias is concerned, in samples with both $T$ and $N$ moderate or large the GMM$pl^{(1)}$ estimator seems preferable over all others considered. When $T$ is large relative to $N$ the GMM$pl^{(2)}$ and GMM$ld^{(2)}$ estimators do not exist and LSDV, GLS and GMM$pl^{(1)}$ seem preferable and when $T$ is small relative to $N$ the GMM$pl$ methods seem preferable to LSDV and GLS. Especially GMM$ld^{(2)}$ does not seem to be recommended in samples where $T$ is large or $\sigma_\eta^2/\sigma_\varepsilon^2$ substantial, but note that in samples with small $T$ its bias may be comparable with that of GMM$ld^{(1)}$ and the GMM$pl$ variants, because these have all a bias of the same order in $N$. Note finally that GMM$pl$ is not invariant regarding $\sigma_\eta^2/\sigma_\varepsilon^2$ and $\pi$ since these affect $Z_l(Z_l'Z_l)^{-1}Z_l'$, and this affects our implementation of LSDVc.

## 5. Bias correction

In principle the bias approximations can be used to construct bias corrected LS and MM estimators. For the LSDV estimator bias correction has been proved reasonably successful in the dynamic model with a strictly exogenous regressor; Kiviet (1995) and Judson and Owen (1999) report a substantial reduction in MSE. However, as can be seen from the bias approximations above, in case of a feedback mechanism in the explanatory variable $x$ the bias depends also on the feedback parameter $\phi$. Appropriate bias correction in this case would require specification and estimation of the model for $x$, which is a major complication. Nevertheless it seems interesting to examine whether bias corrected estimators ignoring the feedback mechanism in $x$ (in practice rather the rule than the exception) may still lead to more efficient estimators in models with lagged feedbacks from $y$ to $x$. Therefore we shall investigate whether the efficiency can be improved by exploiting the bias approximation for LSDV (4.9) upon substituting $\phi = 0$ and $\omega = 0$, when in fact $\phi \neq 0$, i.e. we consider

$$\hat{\delta}_{LSDVc} = \hat{\delta}_{LSDV} - \hat{B}_{LSDV}(T^{-1}), \tag{5.1}$$

with

$$\hat{B}_{LSDV}(T^{-1}) = \hat{\sigma}_\varepsilon^2 \operatorname{tr}(\hat{\Pi})\hat{\Phi}e_{2,1}. \tag{5.2}$$

[1]The difference in order of magnitude of finite sample bias of the GMM$pl$ and GMM$ld$ estimators could be the result of the fact that the latter estimator does not exploit an optimal weighting matrix, see also Alvarez and Arellano (1998).

Note that $\Pi$ is a matrix function of $\gamma$. We will estimate this bias correction using a large-$N$ and large-$T$ consistent estimator for $\gamma$ and $\sigma_\varepsilon^2$ which follows from $\hat{\delta}_{GMMpl}^{(1)}$; for $\hat{\Phi}$ we simply take the observed value $(W'AW)^{-1}$. From the bias approximation (4.9) we can obtain that when $\phi = 0$

$$\mathsf{E}(\hat{\delta}_{LSDVc} - \delta) = O(T^{-1}N^{-1}), \tag{5.3}$$

whereas in case $\phi \neq 0$ this is still $O(T^{-1})$ as for LSDV.

## 6. Simulation results

Data for $y$ have been generated according to equation (2.1) with two different models for the explanatory variable $x$. In scheme 1 its generating equation is designed as in (2.7) with $\bar{x}_{it}$ an AR(1) process, i.e.

$$\left.\begin{aligned} x_{it}^{(1)} &= \bar{x}_{it} + \phi_1 \varepsilon_{i,t-1} + \pi_1 \eta_i \\ \bar{x}_{it} &= \rho_1 \bar{x}_{i,t-1} + \xi_{it} \end{aligned}\right\} \qquad i = 1, ..., N;\ t = 1, ...T, \tag{6.1}$$

where $\xi_{it}$ ~i.i.d.$\mathsf{N}(0, \sigma_\xi^2)$ is independent from $\varepsilon_{it}$ ~i.i.d.$\mathsf{N}(0, \sigma_\varepsilon^2)$, and these are again independent from $\eta_i$ ~i.i.d.$\mathsf{N}(0, \sigma_\eta^2)$. Hence, the explanatory variable $x$ is a linear combination of three stochastically independent components, viz. the AR(1) process $\bar{x}$, the one-period lagged disturbance term $\varepsilon$ and the individual effects $\eta$. It can be rewritten as

$$x_{it}^{(1)} = \frac{1}{1 - \rho_1 L} \xi_{it} + \phi_1 \varepsilon_{i,t-1} + \pi_1 \eta_i, \tag{6.2}$$

where $L$ is the lag operator. As long as $|\rho_1| < 1$ this process will be stationary. When $\phi_1 = 0$ regressor $x$ is strictly exogenous and for $\phi_1 \neq 0$ there is a lagged feedback mechanism in $x$. In the earlier sections this scheme proved to be relatively easy to handle from an analytical point of view. However, from an empirical point of view a more interesting alternative is the process

$$x_{it}^{(2)} = \rho_2 x_{i,t-1}^{(2)} + \phi_2 y_{i,t-1}^{(2)} + \pi_2 \eta_i + \xi_{it}, \tag{6.3}$$

where like in scheme 1 there is lagged feedback, but $x_{it}^{(2)}$ now depends on all past disturbances and not just on $\varepsilon_{i,t-1}$. In this case, indicated as scheme 2, we have

$$x_{it}^{(2)} = \frac{(1 - \gamma L)\xi_{it} + \phi_2 \varepsilon_{i,t-1}}{1 - (\gamma + \beta\phi_2 + \rho_2)L + \gamma\rho_2 L^2} + \frac{(1 - \gamma)\pi_2 + \phi_2}{(1 - \gamma)(1 - \rho_2) - \beta\phi_2} \eta_i. \tag{6.4}$$

Like for scheme 1 this is again a linear combination of three stochastically independent components, but now an ARMA(2,1) process based on $\xi_{it}$, an AR(2) process based on $\varepsilon_{i,t-1}$ and the individual effects. However, in scheme 2 the parameters are not variation free, whereas they are in (6.2): Not just the value of $\phi_2$ itself, but also the values of $\gamma$, $\beta$ and $\rho_2$ affect the extent of the feedback mechanism in $x$ when $\phi_2 \neq 0$. Moreover, the parameters $\gamma$ and $\beta$ determine the $x$ process, whereas they do not under scheme 1. When the $\phi$ coefficients are zero the regressor $x$ is under both schemes a strictly exogenous AR(1) process, but when there is lack of strict exogeneity it is simply ARMA(1,1) under scheme 1 and a much more complicated higher-order ARMA process under scheme 2. For stationarity of $x_{it}^{(2)}$ it is required that the three restrictions

$$\left.\begin{array}{c}\gamma\rho_2 < 1\\ \gamma + \rho_2(1-\gamma) + \beta\phi_2 < 1\\ \gamma + \rho_2(1+\gamma) + \beta\phi_2 > -1\end{array}\right\} \tag{6.5}$$

hold jointly.

With respect to the $x$ processes we chose $\rho_1, \rho_2 = \{0.5, 0.95\}$, and to analyze the impact of lagged feedback under scheme 1 we selected $\phi_1 = \{-1, 0, +1\}$. To maintain some comparability between the results for schemes 1 and 2 (by making the dependence of $x_{it}$ on $\varepsilon_{i,t-1}$ equivalent), we took for $\phi_2$ the values

$$\begin{aligned}\phi_2 &= \phi_1[(1-\gamma)(1-\rho_2) - \beta\phi_2]\\ &= (1-\gamma)(1-\rho_2)\frac{\phi_1}{1+\beta\phi_1}.\end{aligned} \tag{6.6}$$

We fixed $\gamma = 0.75$ and the long-run effect $\theta = \beta/(1-\gamma)$ of $x$ on $y$ has been set equal to unity in all experiments, implying $\beta = 1-\gamma$. This equality and (6.6) reduce the second restriction of (6.5) to $\gamma + (\rho_2 + \phi_2)(1-\gamma) < 1$, which implies that the parameters should obey $\rho_2 + \phi_2 < 1$, i.e.

$$\rho_2 + \frac{\phi_1(1-\gamma)(1-\rho_2)}{1+(1-\gamma)\phi_1} < 1$$

or

$$\frac{\phi_1(1-\gamma)}{1+\phi_1(1-\gamma)} < 1$$

which is true. Hence, $x_t^{(2)}$ is stationary, since it is obvious that our parameter choices also obey the other two restrictions of (6.5). In this study we took $\pi_1 = 0 = \pi_2$. Note that values different from zero would not affect the LSDV estimator, but the finite sample distributions of GLS and GMM estimators (and also bias corrected LSDV based on a preliminary GMM estimate) are affected by correlation between regressors and effects[2].

We shall normalize with respect to the variance of the disturbance term $\sigma_\varepsilon^2$. Hence, the design parameters that remain to be fixed now are $\sigma_\eta^2$ and $\sigma_\xi^2$. Appendix C gives further details on how we attributed values to these two variances. We chose a relationship between $\sigma_\eta^2$ and $\sigma_\varepsilon^2$ such that the impact on $\mathrm{Var}(y_{it})$ of the two variance components $\eta_i$ and $\varepsilon_{it}$ has a ratio $\mu^2$. This implies for scheme 1

$$\sigma_\eta^2 = \mu^2\frac{(1-\gamma)(1+2\gamma\beta\phi_1+\beta^2\phi_1^2)}{(1+\gamma)(1+\beta\pi_1)^2} \tag{6.7}$$

and for scheme 2

$$\begin{aligned}\sigma_\eta^2 &= \mu^2\left(1+\rho_2^2 - 2\rho_2\frac{\gamma+\beta\phi_2+\rho_2}{1+\gamma\rho_2}\right)\left[\frac{1-\rho_2+\beta\pi_2}{(1-\gamma)(1-\rho_2)-\beta\phi_2}\right]^{-2}\times\\ &\quad \left[1-\gamma^2\rho_2^2 - \frac{1-\gamma\rho_2}{1+\gamma\rho_2}(\gamma+\beta\phi_2+\rho_2)^2\right]^{-1}.\end{aligned} \tag{6.8}$$

[2]Not reported simulation results, however, show that the effect on finite sample bias of a non-zero correlation between regressors and individual effects is moderate.

In the simulations we examined $\mu = \{1, 5\}$. The parameter $\sigma_\xi^2$ has been determined by controlling the signal-to-noise ratio (SNR) 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 this basic model characteristic. For our panel model we define for schemes $s = \{1, 2\}$

$$SNR^{(s)} = \frac{\mathrm{Var}(y_{it}^{(s)} - \varepsilon_{it} \mid \eta_i)}{\mathrm{Var}(\varepsilon_{it})} = \frac{\mathrm{Var}(y_{it}^{(s)} \mid \eta_i)}{\mathrm{Var}(\varepsilon_{it})} - 1. \tag{6.9}$$

Normalizing with respect to $\sigma_\varepsilon^2$ this ratio is simply equal to $SNR^{(s)} = \mathrm{Var}(y_{it}^{(s)} \mid \eta_i) - 1$. For both schemes 1 and 2 we choose $SNR^{(s)} = \{3, 9\}$ in the simulations. In Appendix C we show that to achieve this, scheme 1 requires

$$\sigma_\xi^2 = \frac{1}{\beta^2}\left[SNR^{(1)} - \frac{(\gamma + \beta\phi_1)^2}{1 - \gamma^2}\right]\frac{(1 - \gamma^2)(1 - \rho_1^2)(1 - \gamma\rho_1)}{1 + \gamma\rho_1} \tag{6.10}$$

and scheme 2

$$\begin{aligned}\sigma_\xi^2 &= \frac{1}{\beta^2}\left(SNR^{(2)} + 1\right)\left[1 - \gamma^2\rho_2^2 - \frac{1 - \gamma\rho_2}{1 + \gamma\rho_2}(\gamma + \beta\phi_2 + \rho_2)^2\right] \\ &\quad - \frac{1}{\beta^2}\left(1 + \rho_2^2 - 2\rho_2\frac{\gamma + \beta\phi_2 + \rho_2}{1 + \gamma\rho_2}\right).\end{aligned} \tag{6.11}$$

Selfevidently not every combination of values of these parameters is feasible.

The range of values assigned above to the design parameters, i.e. to $\gamma$, $\beta$, $\rho_s$, $\phi_s$, $\pi_s$, $\sigma_\eta^2$ (through $\mu$), $\sigma_\xi^2$ (through SNR) with $\sigma_\varepsilon^2 = 1$, leads to 24 different experiments for each scheme $s = \{1, 2\}$, because $\mu$, $\rho_s$ and $SNR^{(s)}$ obtain only two different values, $\phi_s$ three and the others one. As the theoretical bias approximations show that the effects of dynamic adjustments and lagged feedback are especially important for panels with $T$ small, we focus on this case and choose $T = \{5, 10, 20\}$ and $N = \{20, 50\}$. However, not all 24 parametrizations have been examined for all sample sizes and both schemes.

We analyzed LSDV, GLS, LSDVc, GMM$pl^{(2)}$, GMM$pl^{(1)}$, GMM$ld^{(2)}$ and GMM$ld^{(1)}$. Each Monte Carlo estimate is based on 2500 replications. Table 2 presents and labels a selection of the parametrizations examined. For these selected parametrizations results on the bias, variance and root mean squared error (RMSE) of the various estimators of $\gamma$ and $\beta$ are presented in Tables 3 to 7. Tables 3 and 4 contain results for scheme 1, Tables 5 and 6 for scheme 2 and Table 7 for both schemes because $x$ is strictly exogenous.

In Table 3 several interesting patterns arise from the simulation results, in particular regarding the coefficient estimators for $\gamma$. First, the bias of LSDV and of GMM$pl^{(2)}$ are often substantial and comparable in sign and magnitude, while GMM$pl^{(1)}$ is always less biased. Second, the increase in variance of GMM$pl^{(1)}$ in comparison to GMM$pl^{(2)}$ is substantial, but applying a RMSE criterion the GMM$pl^{(1)}$ estimator is usually more efficient than LSDV and GMM$pl^{(2)}$, and LSDV always more efficient than GMM$pl^{(2)}$. Third, bias correction as proposed in (5.1) usually reduces both bias and RMSE substantially for $\gamma$. Fourth, GLS, GMM$ld^{(2)}$ and GMM$ld^{(1)}$, which all estimate the levels equation, show in general a bias of opposite sign with respect to the methods which estimate a transformed equation and they show relatively low standard errors. Based on a RMSE criterion their accuracy is often close to or occasionally better than LSDVc in these models where $\mu = 1$. Focusing on the effect of the lagged feedback on $x$, we find that negative $\phi$ mitigates and positive $\phi$ aggravates bias and RMSE for LSDV (when $\rho = 0.5$), GLS, GMM*pl* (when $\rho = 0.5$) and GMM*ld*. On LSDVc the value of $\phi$ has erratic effects, possibly due to the dependence on the initial estimate by GMM$pl^{(1)}$.

In Table 4 the effects of increased $N$ on the simulation results of Table 3 have been documented. The same patterns are found. The bias of LSDV and GLS remains more or less the same, while the bias in the GMM estimators has decreased. This reflects the theoretical findings in the previous section, i.e. only the order of the leading bias term for LSDV and GLS does not depend on $N$ while all other estimators are semi-consistent for finite $T$ and $N$ large. Next we turn to the Tables 5 and 6, which show some simulation results for the more realistic scheme 2. In general bias and RMSE are of comparable magnitude as for scheme 1.

Finally, Table 7 presents simulation results for parametrizations where we have varied the signal, the relative magnitude of the error components and the number of time periods $T$. The feedback parameter $\phi$ is set at zero, so schemes 1 and 2 coincide and the regressor $x$ is strictly exogenous. Increasing the signal from 3 to 9 leads on the whole to more accurate estimates. However, increasing $\mu$ from 1 to 5, i.e. increasing the relative effect of $\eta$ on $y$, has marked effects, though not on LSDV, which is invariant in this respect, and little on LSDVc. Especially with respect to $\gamma$ LSDVc performs relatively well when $\mu$ is high. Comparing $T = 5$ and $T = 20$, we find, in correspondence with our theoretical results, that an increase in $T$ may increase the bias of GMM$ld^{(2)}$. We also notice that the bias correction is more successful for higher values of $T$ and, although the bias diminishes for larger $T$, may yield relatively efficient corrected estimates.

## 7. Concluding remarks

In this study we have analyzed the finite sample properties of several least squares and method of moments estimators in panel data models with both a lagged dependent variable regressor and an additional strictly exogenous or predetermined explanatory variable in the presence of unobserved individual effects and white-noise disturbances. The analysis is based on both analytical and experimental methods.

Using asymptotic expansion techniques we have developed bias approximation formulae for several estimators in this model. As can be seen from the leading terms of these bias approximations LSDV and GLS are biased to order $O(T^{-1})$, irrespective of the value of $N$, whereas the MM estimators are biased to order $O(N^{-1})$ assuming $T$ fixed. However, we established more subtle differences between the various MM implementations. We considered two variants of MM estimators using either the levels equation or the model in orthogonal deviations, denoted by GMM*ld* and GMM*pl* respectively. Both have been examined for the situation where all linear moment conditions are being used (which amounts to a number of instruments of order $T^2$) and when fewer instruments are being exploited (we took a number of order $T$). We found that GMM when employing all available linear moment conditions is biased to an order larger in magnitude by a factor $T$ than GMM when using a set of instruments which contain a number of instruments reduced to order $T$. However, this aggravation of bias when the number of instruments is increased from order $T$ to order $T^2$ is not affecting GMM*ld* and GMM*pl* in the same way, because the bias in GMM*pl* is overall of smaller order in $T$ than for GMM*ld*. We expect these differences to be the result of the different weighting matrices used in estimation, see also Alvarez and Arellano (1998). We also found that the leading term in the bias for all estimators indicates that the effects of feedbacks in the explanatory variable (predeterminedness instead of strict exogeneity) is of similar magnitude as the effects of the presence of a lagged dependent variable regressor for that particular estimation technique. Our theoretical derivations also indicate that for the GLS estimator and for the MM estimators of the equation in levels the leading term of the bias is strongly affected by the magnitude of the individual effects, which can be a serious drawback in practice. Whether all these qualitative differences are of practical relevance in actual finite samples cannot be established from this type of analysis, but can by simulation.

In our Monte Carlo experiments we have corroborated most of our theoretical findings. We have put much effort in designing the Monte Carlo such that all aspects that might be relevant are taken care of. However, this implied that a considerable number of design parameters has to be varied. At this stage we did not yet exploit all these dimensions of the design, to limit the number of tables. We examined two different schemes for the feedback mechanism in the explanatory variable. One scheme is very simple and is the one used in the analytical derivations, while the other seems much more relevant from a practitioners point of view. The latter one, although easily handled in a Monte Carlo study, would be difficult to analyze theoretically. However, there are no reasons to suspect that it would lead to different qualitative results regarding the order of magnitude of the leading term in the bias and indeed we find almost similar quantitative results in the simulations.

The actual findings from the Monte Carlo study are as follows. First, they show for all MM estimators a reduction in bias and standard error when $N$ is increased from 20 to 50, while the results for LSDV and GLS remain virtually unchanged. Second, regarding increasing the number of moment conditions used in estimation the simulations show that for GMM estimators (especially those using level instruments) the actual increase in bias is of considerable importance in finite samples. Third, we find that the bias of the GLS and the GMM*ld* estimators, which estimate the untransformed model, depends heavily on the magnitude of the individual effects and may lead to dramatic biases.

It is clear from the simulation results that at particular parameter values all techniques may show substantial distortions in samples where both $T$ and $N$ are moderate. Hence, we conclude that standard first-order asymptotic theory is of little use here. The higher-order asymptotic results developed in this study proved much more informative about the actual finite sample behaviour of the various methods. However, due to the situation that the performance of the techniques is not invariant with respect to parameters which are unknown in practice, a straightforward advise for practitioners regarding which method to prefer in finite samples does not emerge. From the simulations it follows that none of the techniques examined shows superior performance over a wide range of relevant sample dimensions and parameter values, although the simple bias corrected LSDV estimator – which presupposes strict exogeneity – often works remarkably well, especially when the individual effects are a much more prominent error component than the general disturbance term.

**References**

Ahn, S.C., Schmidt, P., 1995. Efficient estimation of models for dynamic panel data. *Journal of Econometrics* 68, 5-27.

Alonso-Borrego, C., Arellano, M., 1999. Symmetrically normalized instrumental-variable estimation using panel data. *Journal of Business & Economic Statistics* 17, 36-49.

Alvarez, J., Arellano, M., 1998. The time series and cross-section asymptotics of dynamic panel data estimators. *CEMFI*, working paper 9808, Madrid.

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

Arellano, M., Bond, S., 1991. Some tests of specification for panel data: Monte Carlo evidence and an application to employment equations. *Review of Economic Studies* 58, 277-297.

Arellano, M., Bover, O., 1995. Another look at the instrumental-variable estimation of error-component models. *Journal of Econometrics* 68, 29-52.

Arellano, M., Honoré, B., 2001. Panel data models: some recent developments. In: Heckman, J.J., Leamer, E. (Eds.), *Handbook of Econometrics Volume 5.* Elsevier Science B.V., The Netherlands (pages 3229-3296).

Blundell, R., Bond, S., 1998. Initial conditions and moment restrictions in dynamic panel data models. *Journal of Econometrics* 87, 115-143.

Blundell, R., Bond, S., Windmeijer, F., 2000. Estimation in dynamic panel data models: improving on the performance of the standard GMM estimators. In: Baltagi, B.H. (Eds.), *Nonstationary Panels, Panel Cointegration, and Dynamic Panels.* Advances in Econometrics 15, Amsterdam: JAI Press, Elsevier Science.

Giersbergen, N.P.A. van, Kiviet, J.F., 1996. Bootstrapping a stable AD model: weak vs strong exogeneity. *Oxford Bulletin of Economics and Statistics* 58, 631-656.

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

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

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

Koenker, R., Machado, J.A.F., 1999. GMM inference when the number of moment conditions is large. *Journal of Econometrics* 93, 327-344.

Ziliak, J.P., 1997. Efficient estimation with panel data when instruments are predetermined: An empirical comparison of moment-condition estimators. *Journal of Business & Economic Statistics* 15, 419-431.

# A. The equivalence of two particular GMM estimators

From formula (3.13) it is immediately clear that the two estimators will be equivalent if $P'Z_l$ and $D'Z_l$ span the same $N(T-1)\times m$ subspace for $Z_l = Z_l^{(2)}$. To verify this we first rearrange the columns of the instrument matrix, so that

$$\begin{aligned} Z_{li}^{*(2)} &= \begin{pmatrix} y_{i0} & 0 & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 & 0 & \cdots & 0 & \cdots & 0 & \cdots \\ 0 & y_{i0} & 0 & \cdots & 0 & y_{i1} & 0 & \cdots & 0 & 0 & \cdots & 0 & & 0 & \\ 0 & 0 & y_{i0} & & 0 & 0 & y_{i1} & & 0 & y_{i2} & & 0 & & \vdots & \\ \vdots & \vdots & & & \vdots & \vdots & & & \vdots & \vdots & & \vdots & \cdots & 0 & \\ 0 & 0 & 0 & \cdots & y_{i0} & 0 & 0 & \cdots & y_{i1} & 0 & & y_{i2} & \cdots & y_{i,T-2} & \cdots \end{pmatrix} \\ &= \left[ y_{i0}I_{T-1}, y_{i1}\begin{pmatrix} 0' \\ I_{T-2}\end{pmatrix}, ..., y_{i,T-2}\begin{pmatrix} 0 \\ 1\end{pmatrix}, x_{i1}I_{T-1}, x_{i2}\begin{pmatrix} 0' \\ I_{T-2}\end{pmatrix}, ..., x_{i,T-1}\begin{pmatrix} 0 \\ 1\end{pmatrix}\right]. \end{aligned}$$

From this, one can see that the equivalence requires: for $s = 1, ..., T-1$ the matrix $P_T'H_s$ spans the same $(T-1)\times s$ subspace as $D_T'H_s$, where $H_s = (O : I_s)'$ is a $(T-1)\times s$ selection matrix. Hence, we have to show that for $s = 1, ..., T-1$ the $s$ last columns of $P_T'$ span the same subspace as the last $s$ columns of $D_T'$.

Note that

$$P_T' = \begin{pmatrix} c_1 & 0 & 0 & \cdots & \cdot & 0 \\ -\frac{c_1}{T-1} & c_2 & 0 & \cdots & \cdot & \cdot \\ -\frac{c_1}{T-1} & -\frac{c_2}{T-2} & c_3 & & & \cdot \\ \vdots & \vdots & & \ddots & & \vdots \\ \cdot & \cdot & & & c_{T-2} & 0 \\ -\frac{c_1}{T-1} & -\frac{c_2}{T-2} & \cdot & \cdots & -\frac{c_{T-2}}{2} & c_{T-1} \\ -\frac{c_1}{T-1} & -\frac{c_2}{T-2} & \cdot & \cdots & -\frac{c_{T-2}}{2} & -\frac{c_{T-1}}{1} \end{pmatrix},$$

with $c_{T-s} = \sqrt{s/(s+1)}$ and

$$D_T' = \begin{pmatrix} -1 & 0 & 0 & \cdots & \cdot & 0 \\ 1 & -1 & 0 & \cdots & \cdot & \cdot \\ 0 & 1 & -1 & & & \cdot \\ \vdots & & & \ddots & & \vdots \\ \cdot & & & & -1 & 0 \\ \cdot & & & & 1 & -1 \\ 0 & 0 & \cdot & \cdots & 0 & 1 \end{pmatrix}.$$

Note further that the final column of $P_T'$ spans the same subspace as the final column of $D_T'$ and that multiplying column $T-s+1$ of $P_T'$ for $s = 2, ..., T-1$ by $\sqrt{(s-1)/(s+1)}$ and subtracting this from column $T-s$ yields a column equal to

$$\left(0, .., 0, c_{T-s}, -\frac{c_{T-s}}{s} - c_{T-s+1}\sqrt{\frac{s-1}{s+1}}, 0, ..., 0\right)',$$

where the final zeros follow from

$$-\frac{c_{T-s}}{s} + \frac{c_{T-s+1}}{s-1}\sqrt{\frac{s-1}{s+1}} = -\sqrt{\frac{1}{s(s+1)}} + \sqrt{\frac{1}{s(s-1)}}\sqrt{\frac{s-1}{s+1}} = 0.$$

Since the other two non-zero elements are respectively $c_{T-s} = \sqrt{s/(s+1)}$ and

$$-\frac{c_{T-s}}{s} - c_{T-s+1}\sqrt{\frac{s-1}{s+1}} = -\sqrt{\frac{1}{s(s+1)}} - \sqrt{\frac{s-1}{s}}\sqrt{\frac{s-1}{s+1}} = -c_{T-s},$$

this column is a multiple of column $T-s$ of $D_T'$, which proves the equivalence.

# B. Asymptotic expansions of estimation errors

The estimation error of the GLS estimator (3.6) is

$$\hat{\delta}_{GLS} - \delta = \left(W'V^{-1}W\right)^{-1} W'V^{-1}u. \tag{B.1}$$

Under the assumptions made in Section 2 we can expand the GLS estimation error in a similar way as has been done for the LSDV estimator in Section 3. Hence, since

$$W'V^{-1}W = [I + (W'V^{-1}W - \Phi_{GLS}^{-1})\Phi_{GLS}]\Phi_{GLS}^{-1}, \tag{B.2}$$

where $\Phi_{GLS}^{-1} = \mathsf{E}(W'V^{-1}W) = O(n)$ and $W'V^{-1}W - \Phi_{GLS}^{-1} = O_p(n^{1/2})$, we find

$$\left(W'V^{-1}W\right)^{-1} = \Phi_{GLS} + O_p(n^{-3/2}). \tag{B.3}$$

Because $W'V^{-1}u = O_p(n^{1/2})$ we obtain for the expected estimation error

$$\mathsf{E}(\hat{\delta}_{GLS} - \delta) = \Phi_{GLS}\mathsf{E}(W'V^{-1}u) + O(n^{-1}). \tag{B.4}$$

The estimation error of the GMM estimator (3.14) can be expressed as

$$\hat{\delta}_{GMMpl} - \delta = (W'P'M_{Z_l}PW)^{-1}W'P'M_{Z_l}P\varepsilon, \tag{B.5}$$

where $M_Z = Z(Z'Z)^{-1}Z'$. From

$$W'P'M_{Z_l}PW = [I + (W'P'M_{Z_l}PW - \Upsilon^{-1})\Upsilon]\Upsilon^{-1}, \tag{B.6}$$

with $\Upsilon^{-1} = \mathsf{E}(W'P'M_{Z_l}PW) = O(n)$, $W'P'M_{Z_l}PW = O_p(n)$ and $W'P'M_{Z_l}PW - \Upsilon^{-1} = O_p(n^{1/2})$ we obtain

$$(W'P'M_{Z_l}PW)^{-1} = \Upsilon + O_p(n^{-3/2}), \tag{B.7}$$

and hence

$$\mathsf{E}(\hat{\delta}_{GMMpl} - \delta) = \Upsilon\mathsf{E}(W'P'M_{Z_l}P\varepsilon) + O(n^{-1}),$$

because $W'P'M_{Z_l}P\varepsilon = O_p(n^{1/2})$.

Using similar reasoning, the bias of the GMM estimator (3.19) is

$$\mathsf{E}(\hat{\delta}_{GMMld} - \delta) = \Theta\mathsf{E}(W'J'M_{Z_d}Ju) + O(n^{-1}), \tag{B.8}$$

with $\Theta^{-1} = \mathsf{E}(W'J'M_{Z_d}JW)$.

## C. Derivation of the bias approximations

Here we will evaluate the leading terms in (4.4) through (4.7). From (2.16) it follows that

$$\begin{aligned} \tilde{y}_{(-1)} &= L\tilde{y} + (I_N \otimes e_{T,1})\tilde{y}_0 \qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad \text{(C.1)} \\ &= \omega(I_N \otimes e_{T,1})\varepsilon_0 + L\Gamma(I_{NT} + \beta\phi L)\varepsilon + (\omega\gamma + \phi\beta)L\Gamma(I_N \otimes e_{T,1})\varepsilon_0 + \alpha S\eta, \end{aligned}$$

where $\tilde{y}_0 = (\tilde{y}_{10}, ..., \tilde{y}_{N0})'$ has been defined in (2.10). With (2.17) this gives

$$\begin{aligned} \tilde{W} &= (\tilde{y}_{(-1)}, \tilde{x}) \\ &= S\eta \left(\alpha e_{2,1}' + \pi e_{2,2}'\right) + L\Gamma(I_{NT} + \beta\phi L)\varepsilon e_{2,1}' + \phi L\varepsilon e_{2,2}' + \omega(I_N \otimes e_{T,1})\varepsilon_0 e_{2,1}' \\ &\quad + (\omega\gamma + \phi\beta)L\Gamma(I_N \otimes e_{T,1})\varepsilon_0 e_{2,1}' + \phi(I_N \otimes e_{T,1})\varepsilon_0 e_{2,2}'. \qquad \text{(C.2)} \end{aligned}$$

We first analyze the expected LSDV estimation error in (4.4). Since

$$\mathsf{E}(W'A\varepsilon) = \mathsf{E}(\tilde{W}'A\varepsilon) + \mathsf{E}(\bar{W}'A\varepsilon) = \mathsf{E}(\tilde{W}'A\varepsilon), \qquad \text{(C.3)}$$

noting that $\mathsf{E}(\bar{W}'A\varepsilon) = 0$, and using $\Pi = AL\Gamma$, we find

$$\begin{aligned} \mathsf{E}(W'A\varepsilon) &= \mathsf{E}[\varepsilon'\Pi(I_{NT} + \beta\phi L)\varepsilon]e_{2,1} + \phi\mathsf{E}(\varepsilon' AL\varepsilon)e_{2,2} \\ &= \sigma_\varepsilon^2 \operatorname{tr}(\Pi)e_{2,1} + \sigma_\varepsilon^2\beta\phi \operatorname{tr}(\Pi L)e_{2,1} + \sigma_\varepsilon^2\phi \operatorname{tr}(AL)e_{2,2} \\ &= O(N), \qquad \text{(C.4)} \end{aligned}$$

because $\varepsilon_0$ and $\varepsilon$ are independent by assumption. Substituting (C.4) in (4.4) leads to (4.8) with (4.9).

Regarding the GLS estimation error in (4.5) we obtain

$$\begin{aligned} \mathsf{E}(W'V^{-1}u) &= \mathsf{E}(\tilde{W}'V^{-1}u) + \mathsf{E}(\bar{W}'V^{-1}u) \\ &= \mathsf{E}(\tilde{W}'V^{-1}S\eta) + \mathsf{E}(\tilde{W}'V^{-1}\varepsilon). \qquad \text{(C.5)} \end{aligned}$$

Employing (C.2) we find

$$\begin{aligned} \mathsf{E}(\tilde{W}'V^{-1}S\eta) &= \mathsf{E}(\eta' S' V^{-1} S\eta)(\alpha e_{2,1} + \pi e_{2,2}) \\ &= \sigma_\eta^2 \operatorname{tr}\left(S'V^{-1}S\right)(\alpha e_{2,1} + \pi e_{2,2}) \\ &= \sigma_\eta^2 \frac{NT}{1 + T\frac{\sigma_\eta^2}{\sigma_\varepsilon^2}}(\alpha e_{2,1} + \pi e_{2,2}). \qquad \text{(C.6)} \end{aligned}$$

In a similar way as in (C.4)

$$\mathsf{E}(\tilde{W}'V^{-1}\varepsilon) = \sigma_\varepsilon^2 \operatorname{tr}(V^{-1}L\Gamma)e_{2,1} + \sigma_\varepsilon^2\beta\phi \operatorname{tr}(V^{-1}L\Gamma L)e_{2,1} + \sigma_\varepsilon^2\phi \operatorname{tr}(V^{-1}L)e_{2,2}. \qquad \text{(C.7)}$$

Both (C.6) and (C.7) are $O(N)$ and therefore lead to (4.10).

Next, we consider for $r = \{1, 2\}$ the expected GMM$pl^{(r)}$ estimation error (4.6). We examine

$$\mathsf{E}(W'P'M_{Z_l}P\varepsilon) = \sum_{t=1}^{T-1} \mathsf{E}(W_t^{*\prime} M_{Z_{lt}}\varepsilon_t^*), \qquad \text{(C.8)}$$

where $M_{Z_{lt}} = Z_{lt}(Z_{lt}'Z_{lt})^{-1}Z_{lt}'$ with $\operatorname{tr}(M_{Z_{lt}}) = m_{lt}^{(r)}$. We proceed as follows, see also Alvarez and Arellano (1998). We have

$$\begin{aligned}
\mathsf{E}(W_t^{*\prime}M_{Z_{lt}}\varepsilon_t^*) &= \mathsf{E}\begin{pmatrix} y_{t-1}^{*\prime}M_{Z_{lt}}\varepsilon_t^* \\ x_t^{*\prime}M_{Z_{lt}}\varepsilon_t^* \end{pmatrix} = \mathsf{E}\begin{pmatrix} \operatorname{tr}(y_{t-1}^{*\prime}M_{Z_{lt}}\varepsilon_t^*) \\ \operatorname{tr}(x_t^{*\prime}M_{Z_{lt}}\varepsilon_t^*) \end{pmatrix} \\
&= \mathsf{E}\mathsf{E}_{t-1}\begin{pmatrix} \operatorname{tr}(M_{Z_{lt}}\varepsilon_t^* y_{t-1}^{*\prime}) \\ \operatorname{tr}(M_{Z_{lt}}\varepsilon_t^* x_t^{*\prime}) \end{pmatrix},
\end{aligned}$$

where $\mathsf{E}_{t-1}$ indicates the expectation conditional on information up to $t-1$. Note that $Z_{lt}$ contains only relevant stochastic elements that have been observed prior to $t$, hence

$$\mathsf{E}\mathsf{E}_{t-1}\begin{pmatrix} \operatorname{tr}(M_{Z_{lt}}\varepsilon_t^* y_{t-1}^{*\prime}) \\ \operatorname{tr}(M_{Z_{lt}}\varepsilon_t^* x_t^{*\prime}) \end{pmatrix} = \begin{pmatrix} \mathsf{E}\{\operatorname{tr}[M_{Z_{lt}}\mathsf{E}_{t-1}(\varepsilon_t^* y_{t-1}^{*\prime})]\} \\ \mathsf{E}\{\operatorname{tr}[M_{Z_{lt}}\mathsf{E}_{t-1}(\varepsilon_t^* x_t^{*\prime})]\} \end{pmatrix}.$$

Since $\varepsilon_t^*$ depends exclusively on disturbances drawn since $t$ and has expectation zero it has zero covariance with all components of $W_t^*$ that were observed prior to $t$. That means that conditioning is inconsequential here, i.e.

$$\begin{gathered}
\mathsf{E}_{t-1}(\varepsilon_t^* y_{t-1}^{*\prime}) = \mathsf{E}(\varepsilon_t^* y_{t-1}^{*\prime}) = I_N\mathsf{E}(\varepsilon_{it}^* y_{i,t-1}^*) \\
\mathsf{E}_{t-1}(\varepsilon_t^* x_t^{*\prime}) = \mathsf{E}(\varepsilon_t^* x_t^{*\prime}) = I_N\mathsf{E}(\varepsilon_{it}^* x_{it}^*),
\end{gathered}$$

where use has been made of the independence of the $N$ individuals. The above implies

$$\begin{aligned}
\mathsf{E}(W_t^{*\prime}M_{Z_{lt}}\varepsilon_t^*) &= m_{lt}^{(r)}\begin{pmatrix} \mathsf{E}(\varepsilon_{it}^* y_{i,t-1}^*) \\ \mathsf{E}(\varepsilon_{it}^* x_{it}^*) \end{pmatrix} = m_{lt}^{(r)}\begin{pmatrix} \mathsf{E}(p_t'\varepsilon_i p_t' y_{i(-1)}) \\ \mathsf{E}(p_t'\varepsilon_i p_t' x_i) \end{pmatrix} \\
&= m_{lt}^{(r)}\begin{pmatrix} \operatorname{tr}[\mathsf{E}(\varepsilon_i y_{i(-1)}')p_t p_t'] \\ \operatorname{tr}[\mathsf{E}(\varepsilon_i x_i')p_t p_t'] \end{pmatrix},
\end{aligned}$$

where $p_t'$ is the $t^{th}$ row of the matrix $P_T$. Hence,

$$\sum_{t=1}^{T-1}\mathsf{E}(W_t^{*\prime}M_{Z_{lt}}\varepsilon_t^*) = \begin{pmatrix} \operatorname{tr}[\mathsf{E}(\varepsilon_i y_{i(-1)}')\sum_{t=1}^{T-1} m_{lt}^{(r)}p_t p_t'] \\ \operatorname{tr}[\mathsf{E}(\varepsilon_i x_i')\sum_{t=1}^{T-1} m_{lt}^{(r)}p_t p_t'] \end{pmatrix}. \tag{C.9}$$

When using all available moment conditions, i.e. $Z_l = Z_l^{(2)}$, we have $m_{lt}^{(2)} = 2t$, hence (see again Alvarez and Arellano, 1998)

$$\sum_{t=1}^{T-1} m_{lt}^{(r)}p_t p_t' = 2\sum_{t=1}^{T-1} t p_t p_t' = 2\sum_{s=2}^{T} H_s A_s H_s' \tag{C.10}$$

where $H_s = (O : I_s)'$ is a $T \times s$ selection matrix and $A_s = I_s - \frac{1}{s}\iota_s\iota_s'$. We examine the second element of (C.9) first. Using (2.12) we have

$$\begin{aligned}
\operatorname{tr}[\mathsf{E}(\varepsilon_i x_i')\sum_{t=1}^{T-1} m_{lt}^{(2)}p_t p_t'] &= 2\sum_{s=2}^{T}\mathsf{E}(\tilde{x}_i' H_s A_s H_s'\varepsilon_i) = 2\phi\sum_{s=2}^{T}\mathsf{E}(\varepsilon_i' L_T' H_s A_s H_s'\varepsilon_i) \\
&= 2\sigma_\varepsilon^2\phi\sum_{s=2}^{T}\operatorname{tr}(H_s A_s H_s' L_T) = O(T),
\end{aligned} \tag{C.11}$$

because $H_sA_sH_s'L_T$ has $s-1$ diagonal elements $-\frac{1}{s}$ and all others 0, thus its trace is $O(1)$ and summing yields $O(T)$. For the first element of (C.9) we find, using (2.14) and noting that $\mathsf{E}(\bar{y}_{i(-1)}'H_sA_sH_s'\varepsilon_i)=0$ and $\varepsilon$ is uncorrelated with $\eta$ and $\varepsilon_0$,

$$\begin{aligned}\operatorname{tr}[\mathsf{E}(\varepsilon_i y_{i(-1)}')\sum_{t=1}^{T-1} m_{lt}^{(r)} p_t p_t'] &= 2\sum_{s=2}^{T}\mathsf{E}(\tilde{y}_{i(-1)}'H_sA_sH_s'\varepsilon_i) \\ &= 2\sum_{s=2}^{T}\mathsf{E}\left[\varepsilon_i'(I_T+\beta\phi L_T')\Gamma_T'L_T'H_sA_sH_s'\varepsilon_i\right] \\ &= 2\sigma_\varepsilon^2\sum_{s=2}^{T}\operatorname{tr}[H_sA_sH_s'L_T\Gamma_T(I_T+\beta\phi L_T)]=O(T).\end{aligned} \tag{C.12}$$

Hence, we obtain

$$\mathsf{E}(W'P'M_{Z_l^{(2)}}P\varepsilon)=O(T). \tag{C.13}$$

In the second case, i.e. $m_{lt}^{(1)}=\operatorname{tr}(M_{Z_{lt}^{(1)}})=2$, we have

$$\sum_{t=1}^{T-1} m_{lt}^{(1)} p_t p_t' = 2\sum_{t=1}^{T-1} p_t p_t' = 2A_T, \tag{C.14}$$

so

$$\begin{aligned}\operatorname{tr}[\mathsf{E}(\varepsilon_i y_{i(-1)}')\sum_{t=1}^{T-1} m_{lt}^{(1)} p_t p_t'] &= 2\mathsf{E}(\tilde{y}_{i(-1)}'A_T\varepsilon_i) \\ &= 2\sigma_\varepsilon^2[\operatorname{tr}(\Pi_T)+\beta\phi\operatorname{tr}(\Pi_T L_T)]=O(1)\end{aligned} \tag{C.15}$$

and

$$\begin{aligned}\operatorname{tr}[\mathsf{E}(\varepsilon_i x_i')\sum_{t=1}^{T-1} m_{lt}^{(1)} p_t p_t'] &= 2\mathsf{E}(\tilde{x}_i'A_T\varepsilon_i) \\ &= 2\sigma_\varepsilon^2\phi\operatorname{tr}(A_T L_T)=O(1),\end{aligned} \tag{C.16}$$

giving

$$\mathsf{E}(W'P'M_{Z_l^{(1)}}P\varepsilon)=O(1). \tag{C.17}$$

Combining expressions (C.13) and (C.17) with (4.6) leads to the bias approximations (4.13) and (4.15).

Finally, we consider (4.7) for the GMM estimator using instruments in first differences and examine

$$\mathsf{E}(W'J'M_{Z_d}Ju)=\sum_{t=2}^{T}\mathsf{E}(W_t'M_{Z_{dt}}u_t), \tag{C.18}$$

where $\operatorname{tr}(M_{Z_{dt}})=m_{dt}^{(r)}$ is again either $2t$ or 2. We have

$$\mathsf{E}(W_t'M_{Z_{dt}}u_t)=\mathsf{E}\begin{pmatrix} y_{t-1}'M_{Z_{dt}}u_t \\ x_t'M_{Z_{dt}}u_t\end{pmatrix}=\mathsf{E}\begin{pmatrix} y_{t-1}'M_{Z_{dt}}\eta \\ x_t'M_{Z_{dt}}\eta\end{pmatrix}+\mathsf{E}\begin{pmatrix} y_{t-1}'M_{Z_{dt}}\varepsilon_t \\ x_t'M_{Z_{dt}}\varepsilon_t\end{pmatrix},$$

where the last term equals zero, because all stochastic elements of $y_{t-1}$, $x_t$ and $Z_{dt}$ are independent of $\varepsilon_t$. For the initial term we have

$$\mathsf{E}\begin{pmatrix} y'_{t-1}M_{Z_{dt}}\eta \\ x'_t M_{Z_{dt}}\eta \end{pmatrix} = \mathsf{E}\begin{pmatrix} \tilde{y}'_{t-1}M_{Z_{dt}}\eta \\ \tilde{x}'_t M_{Z_{dt}}\eta \end{pmatrix} + \mathsf{E}\begin{pmatrix} \bar{y}'_{t-1}M_{Z_{dt}}\eta \\ \bar{x}'_t M_{Z_{dt}}\eta \end{pmatrix},$$

where the last term is again zero, now because $\bar{y}_{t-1}$, $\bar{x}_t$ and $Z_{dt}$ are independent of $\eta$. For the initial term we find, by removing all random elements from $\tilde{y}_{t-1}$ and $\tilde{x}_t$ that are also independent of $\eta$ and therefore cannot contribute to the expectation,

$$\mathsf{E}\begin{pmatrix} \tilde{y}'_{t-1}M_{Z_{dt}}\eta \\ \tilde{x}'_t M_{Z_{dt}}\eta \end{pmatrix} = \mathsf{E}\begin{pmatrix} \alpha\eta' M_{Z_{dt}}\eta \\ \pi\eta' M_{Z_{dt}}\eta \end{pmatrix} = \mathsf{E}\,\mathrm{tr}(\eta' M_{Z_{dt}}\eta)\begin{pmatrix} \alpha \\ \pi \end{pmatrix}.$$

Because

$$\mathsf{E}\,\mathrm{tr}(\eta' M_{Z_{dt}}\eta) = \mathsf{E}\,\mathrm{tr}(M_{Z_{dt}}\eta\eta') = \mathsf{E}\,\mathrm{tr}[M_{Z_{dt}}\mathsf{E}(\eta\eta')] = \sigma_\eta^2\,\mathrm{tr}(M_{Z_{dt}}) = \sigma_\eta^2 m_{dt}^{(r)},$$

we find now

$$\mathsf{E}(W'J'M_{Z_d}Ju) = \sigma_\eta^2 \sum_{t=2}^{T} m_{dt}^{(r)} \begin{pmatrix} \alpha \\ \pi \end{pmatrix}, \tag{C.19}$$

which yields for $Z_d = Z_d^{(2)}$, i.e. using all instruments,

$$\mathsf{E}(W'J'M_{Z_d^{(2)}}Ju) = \sigma_\eta^2(T^2+T-2)\begin{pmatrix} \alpha \\ \pi \end{pmatrix} = O(T^2),$$

so that result (4.17) follows. For $Z_d = Z_d^{(1)}$ we find

$$\mathsf{E}(W'J'M_{Z_d^{(1)}}Ju) = 2\sigma_\eta^2(T-1)\begin{pmatrix} \alpha \\ \pi \end{pmatrix} = O(T),$$

leading to the expression in (4.19).

# D. Details on the simulation design

In what follows we make use of some standard results for stationary AR(2) processes. Let

$$z_t = \frac{1}{1-\alpha_1 L - \alpha_2 L^2}\varepsilon_t, \tag{D.1}$$

where $\varepsilon_t \sim \text{i.i.d.}(0, \sigma_\varepsilon^2)$, then

$$\mathrm{Var}(z_t) = \left(1 - \alpha_2^2 - \frac{1+\alpha_2}{1-\alpha_2}\alpha_1^2\right)^{-1}\sigma_\varepsilon^2 \tag{D.2}$$

and

$$\mathrm{Cov}(z_t, z_{t-1}) = \frac{\alpha_1}{1-\alpha_2}\,\mathrm{Var}(z_t). \tag{D.3}$$

Hence, for the ARMA(2,1) process

$$w_t = (1+\theta L)z_t = \frac{1+\theta L}{1-\alpha_1 L - \alpha_2 L^2}\varepsilon_t, \tag{D.4}$$

we find

$$\begin{aligned} \mathrm{Var}(w_t) &= (1+\theta^2)\,\mathrm{Var}(z_t) + 2\theta\,\mathrm{Cov}(z_t, z_{t-1}) \\ &= \left(1+\theta^2+2\theta\frac{\alpha_1}{1-\alpha_2}\right)\mathrm{Var}(z_t). \end{aligned} \tag{D.5}$$

Note that in model (2.1)

$$y_{it} = \frac{\beta}{1-\gamma L}x_{it} + \frac{1}{1-\gamma}\eta_i + \frac{1}{1-\gamma L}\varepsilon_{it}. \tag{D.6}$$

Substitution of (6.2) gives for scheme 1

$$y_{it}^{(1)} = \frac{\beta}{(1-\gamma L)(1-\rho_1 L)}\xi_{it} + \frac{1+\beta\pi_1}{1-\gamma}\eta_i + \frac{1+\beta\phi_1 L}{1-\gamma L}\varepsilon_{it}. \tag{D.7}$$

Hence, using (D.2) and (D.5), we find

$$\begin{aligned} \mathrm{Var}(y_{it}^{(1)}) &= \sigma_\xi^2\beta^2\frac{1+\gamma\rho_1}{(1-\rho_1^2)(1-\gamma^2)(1-\gamma\rho_1)} \\ &\quad + \sigma_\varepsilon^2\frac{1+2\gamma\beta\phi_1+\beta^2\phi_1^2}{1-\gamma^2} + \sigma_\eta^2\left(\frac{1+\beta\pi_1}{1-\gamma}\right)^2. \end{aligned} \tag{D.8}$$

This yields

$$SNR^{(1)} = \sigma_\xi^2\beta^2\frac{1+\gamma\rho_1}{(1-\rho_1^2)(1-\gamma^2)(1-\gamma\rho_1)} + \frac{(\gamma+\beta\phi_1)^2}{1-\gamma^2} \tag{D.9}$$

from which (6.10) easily follows. From (D.8) it is also found that, in order to achieve contributions to $\mathrm{Var}[y_{it}^{(1)}]$ from the variance components $\eta_i$ and $\varepsilon_{it}$ in a proportion with ratio $\mu^2$, one has to choose

$$\sigma_\eta^2\left(\frac{1+\beta\pi_1}{1-\gamma}\right)^2 = \mu^2\sigma_\varepsilon^2\left(\frac{1+2\gamma\beta\phi_1+\beta^2\phi_1^2}{1-\gamma^2}\right), \tag{D.10}$$

from which (6.7) follows.

Model (2.1) can also be rewritten as

$$(1-\rho_2 L)(1-\gamma L)y_{it} = \beta(1-\rho_2 L)x_{it} + (1-\rho_2)\eta_i + (1-\rho_2 L)\varepsilon_{it}. \tag{D.11}$$

According to (6.4) we have in scheme 2

$$(1-\rho_2 L)x_{it}^{(2)} = \phi_2 L y_{it}^{(2)} + \pi_2\eta_i + \xi_{it}. \tag{D.12}$$

Substitution in (D.11) yields

$$(1-\rho_2 L)(1-\gamma L)y_{it}^{(2)} = \beta\phi_2 L y_{it}^{(2)} + \beta\xi_{it} + (1-\rho_2+\beta\pi_2)\eta_i + (1-\rho_2 L)\varepsilon_{it}$$

or

$$(1-\alpha_1 L-\alpha_2 L^2)y_{it}^{(2)}=\beta\xi_{it}+(1-\rho_2+\beta\pi_2)\eta_i+(1-\rho_2 L)\varepsilon_{it},$$

with

$$\left.\begin{aligned}\alpha_1&=\gamma+\beta\phi_2+\rho_2\\ \alpha_2&=-\gamma\rho_2.\end{aligned}\right\} \tag{D.13}$$

Hence,

$$y_{it}^{(2)}=\frac{1}{1-\alpha_1 L-\alpha_2 L^2}[\beta\xi_{it}+(1-\rho_2+\beta\pi_2)\eta_i+(1-\rho_2 L)\varepsilon_{it}] \tag{D.14}$$

from which we obtain

$$\begin{aligned}\mathrm{Var}[y_{it}^{(2)}] &= \left[\sigma_\xi^2\beta^2+\sigma_\varepsilon^2\left(1+\rho_2^2-2\rho_2\frac{\gamma+\beta\phi_2+\rho_2}{1+\gamma\rho_2}\right)\right]\times\\ &\qquad\left[1-\gamma^2\rho_2^2-\frac{1-\gamma\rho_2}{1+\gamma\rho_2}(\gamma+\beta\phi_2+\rho_2)^2\right]^{-1}\\ &\quad+\sigma_\eta^2\left[\frac{1-\rho_2+\beta\pi_2}{(1-\gamma)(1-\rho_2)-\beta\phi_2}\right]^2.\end{aligned} \tag{D.15}$$

This yields

$$\begin{aligned}SNR^{(2)} &= \left[\sigma_\xi^2\beta^2+\sigma_\varepsilon^2\left(1+\rho_2^2-2\rho_2\frac{\gamma+\beta\phi_2+\rho_2}{1+\gamma\rho_2}\right)\right]\times\\ &\qquad\left[1-\gamma^2\rho_2^2-\frac{1-\gamma\rho_2}{1+\gamma\rho_2}(\gamma+\beta\phi_2+\rho_2)^2\right]^{-1}-1,\end{aligned} \tag{D.16}$$

from which (6.11) follows. A similar derivation for a model in a non-panel data context with an integrated weakly exogenous explanatory variable can be found in van Giersbergen and Kiviet (1996). To control the ratio of the variance components in scheme 2 we set

$$\begin{aligned}\sigma_\eta^2\left[\frac{1-\rho_2+\beta\pi_2}{(1-\gamma)(1-\rho_2)-\beta\phi_2}\right]^2 &= \mu^2\sigma_\varepsilon^2\left(1+\rho_2^2-2\rho_2\frac{\gamma+\beta\phi_2+\rho_2}{1+\gamma\rho_2}\right)\times\\ &\qquad\left[1-\gamma^2\rho_2^2-\frac{1-\gamma\rho_2}{1+\gamma\rho_2}(\gamma+\beta\phi_2+\rho_2)^2\right]^{-1},\end{aligned} \tag{D.17}$$

which yields (6.8).

The data have been generated by setting $x_{i,-49}=y_{i,-49}=0$, generating $T+50$ observations and discarding the first 50 observations to minimize the effects of the initial zero values.

Table 1: Characteristics and order of magnitude of finite sample bias

| | dimensional restrictions | equation formulated in: levels | orthogonal deviations | first differences | bias affected by $\sigma^2_\eta/\sigma^2_\varepsilon$ and $\pi$ |
|---|---|---|---|---|---|
| $LSDV$ | – | – | $O(T^{-1})$ | – | no |
| $GLS$ | – | $O(T^{-1})$ | – | $O(T^{-1})$ | heavily |
| $GMMpl^{(2)}$ | $N \geq T-1$ | – | $O(N^{-1})$ | $O(N^{-1})$ | yes |
| $GMMpl^{(1)}$ | – | – | $O(N^{-1}T^{-1})$ | – | yes |
| $GMMld^{(2)}$ | $N \geq T$ | $O(TN^{-1})$ | – | – | heavily |
| $GMMld^{(1)}$ | – | $O(N^{-1})$ | – | – | heavily |

Table 2: Parameter combinations for simulations

| | $s$ | $\gamma$ | $\phi_s$ | $\pi_s$ | $\rho_s$ | $\sigma_\varepsilon$ | $\mu$ | $SNR^{(s)}$ | $\sigma_\eta$ | $\sigma_\xi$ |
|---|---|---|---|---|---|---|---|---|---|---|
| I | 1 | 0.75 | -1 | 0 | 0.5 | 1 | 1 | 3 | 0.31 | 2.41 |
| II | 1 | 0.75 | 0 | 0 | 0.5 | 1 | 1 | 3 | 0.38 | 2.02 |
| III | 1 | 0.75 | 1 | 0 | 0.5 | 1 | 1 | 3 | 0.45 | 1.31 |
| IV | 1 | 0.75 | -1 | 0 | 0.95 | 1 | 1 | 3 | 0.31 | 0.53 |
| V | 1 | 0.75 | 0 | 0 | 0.95 | 1 | 1 | 3 | 0.38 | 0.44 |
| VI | 1 | 0.75 | 1 | 0 | 0.95 | 1 | 1 | 3 | 0.45 | 0.29 |
| VII | 2 | 0.75 | -0.17 | 0 | 0.5 | 1 | 1 | 3 | 0.46 | 2.57 |
| VIII | 2 | 0.75 | 0 | 0 | 0.5 | 1 | 1 | 3 | 0.38 | 2.02 |
| IX | 2 | 0.75 | 0.10 | 0 | 0.5 | 1 | 1 | 3 | 0.33 | 1.60 |
| X | 2 | 0.75 | -0.02 | 0 | 0.95 | 1 | 1 | 3 | 0.49 | 0.52 |
| XI | 2 | 0.75 | 0 | 0 | 0.95 | 1 | 1 | 3 | 0.38 | 0.44 |
| XII | 2 | 0.75 | 0.01 | 0 | 0.95 | 1 | 1 | 3 | 0.31 | 0.39 |
| XIII | 1,2 | 0.75 | 0 | 0 | 0.5 | 1 | 1 | 3 | 0.38 | 2.02 |
| XIV | 1,2 | 0.75 | 0 | 0 | 0.5 | 1 | 1 | 9 | 0.38 | 4.29 |
| XV | 1,2 | 0.75 | 0 | 0 | 0.5 | 1 | 5 | 3 | 1.89 | 2.02 |
| XVI | 1,2 | 0.75 | 0 | 0 | 0.5 | 1 | 5 | 9 | 1.89 | 4.29 |

Table 3: Simulation results for scheme 1, $T = 10$ and $N = 20$

| | | $Bias\ \gamma$ | $Bias\ \beta$ | $std\ \gamma$ | $std\beta$ | $RMSE\ \gamma$ | $RMSE\ \beta$ |
|---|---|---|---|---|---|---|---|
| I | $LSDV$ | -0.11 | 0.00 | 0.06 | 0.03 | 0.12 | 0.03 |
| | $LSDVc$ | 0.01 | 0.02 | 0.07 | 0.03 | 0.07 | 0.03 |
| | $GLS$ | 0.05 | 0.00 | 0.04 | 0.03 | 0.06 | 0.03 |
| | $GMMpl^{(2)}$ | -0.11 | -0.01 | 0.08 | 0.04 | 0.13 | 0.04 |
| | $GMMpl^{(1)}$ | -0.04 | -0.01 | 0.10 | 0.04 | 0.11 | 0.04 |
| | $GMMld^{(2)}$ | 0.05 | 0.00 | 0.05 | 0.03 | 0.07 | 0.03 |
| | $GMMld^{(1)}$ | 0.04 | 0.00 | 0.08 | 0.04 | 0.09 | 0.04 |
| II | $LSDV$ | -0.14 | 0.00 | 0.06 | 0.04 | 0.15 | 0.04 |
| | $LSDVc$ | -0.03 | 0.00 | 0.07 | 0.04 | 0.07 | 0.04 |
| | $GLS$ | 0.07 | -0.01 | 0.04 | 0.03 | 0.08 | 0.04 |
| | $GMMpl^{(2)}$ | -0.14 | -0.01 | 0.08 | 0.04 | 0.16 | 0.04 |
| | $GMMpl^{(1)}$ | -0.05 | -0.01 | 0.10 | 0.05 | 0.11 | 0.05 |
| | $GMMld^{(2)}$ | 0.07 | -0.01 | 0.05 | 0.04 | 0.08 | 0.04 |
| | $GMMld^{(1)}$ | 0.05 | -0.00 | 0.08 | 0.05 | 0.09 | 0.05 |
| III | $LSDV$ | -0.17 | 0.01 | 0.06 | 0.05 | 0.18 | 0.05 |
| | $LSDVc$ | -0.05 | -0.02 | 0.07 | 0.05 | 0.09 | 0.05 |
| | $GLS$ | 0.11 | -0.05 | 0.04 | 0.04 | 0.12 | 0.07 |
| | $GMMpl^{(2)}$ | -0.18 | -0.01 | 0.08 | 0.05 | 0.20 | 0.05 |
| | $GMMpl^{(1)}$ | -0.07 | -0.00 | 0.12 | 0.05 | 0.14 | 0.05 |
| | $GMMld^{(2)}$ | 0.10 | -0.04 | 0.05 | 0.05 | 0.11 | 0.06 |
| | $GMMld^{(1)}$ | 0.07 | -0.02 | 0.08 | 0.06 | 0.11 | 0.06 |
| IV | $LSDV$ | -0.20 | -0.03 | 0.09 | 0.08 | 0.22 | 0.08 |
| | $LSDVc$ | 0.06 | 0.11 | 0.21 | 0.13 | 0.22 | 0.17 |
| | $GLS$ | 0.06 | -0.01 | 0.04 | 0.05 | 0.07 | 0.05 |
| | $GMMpl^{(2)}$ | -0.23 | -0.06 | 0.14 | 0.11 | 0.27 | 0.13 |
| | $GMMpl^{(1)}$ | -0.15 | -0.08 | 0.26 | 0.21 | 0.29 | 0.23 |
| | $GMMld^{(2)}$ | 0.05 | -0.00 | 0.05 | 0.06 | 0.07 | 0.06 |
| | $GMMld^{(1)}$ | 0.04 | 0.01 | 0.09 | 0.09 | 0.10 | 0.09 |
| V | $LSDV$ | -0.21 | 0.06 | 0.07 | 0.14 | 0.22 | 0.16 |
| | $LSDVc$ | -0.05 | 0.02 | 0.09 | 0.13 | 0.10 | 0.13 |
| | $GLS$ | 0.11 | -0.09 | 0.05 | 0.07 | 0.12 | 0.11 |
| | $GMMpl^{(2)}$ | -0.22 | 0.03 | 0.10 | 0.22 | 0.24 | 0.22 |
| | $GMMpl^{(1)}$ | -0.10 | -0.01 | 0.15 | 0.40 | 0.18 | 0.40 |
| | $GMMld^{(2)}$ | 0.08 | -0.06 | 0.05 | 0.09 | 0.10 | 0.11 |
| | $GMMld^{(1)}$ | 0.06 | -0.04 | 0.09 | 0.18 | 0.11 | 0.19 |
| VI | $LSDV$ | -0.21 | 0.05 | 0.07 | 0.08 | 0.22 | 0.09 |
| | $LSDVc$ | -0.05 | -0.06 | 0.10 | 0.10 | 0.11 | 0.12 |
| | $GLS$ | 0.14 | -0.13 | 0.04 | 0.06 | 0.14 | 0.14 |
| | $GMMpl^{(2)}$ | -0.23 | 0.02 | 0.10 | 0.09 | 0.25 | 0.09 |
| | $GMMpl^{(1)}$ | -0.10 | 0.01 | 0.15 | 0.11 | 0.18 | 0.11 |
| | $GMMld^{(2)}$ | 0.12 | -0.10 | 0.05 | 0.07 | 0.13 | 0.12 |
| | $GMMld^{(1)}$ | 0.08 | -0.05 | 0.09 | 0.09 | 0.12 | 0.10 |

Table 4: Simulation results for scheme 1, $T = 10$ and $N = 50$

| | | $Bias\ \gamma$ | $Bias\ \beta$ | $std\ \gamma$ | $std\beta$ | $RMSE\ \gamma$ | $RMSE\ \beta$ |
|---|---|---|---|---|---|---|---|
| I | $LSDV$ | -0.11 | 0.00 | 0.04 | 0.02 | 0.11 | 0.02 |
| | $LSDVc$ | 0.02 | 0.02 | 0.04 | 0.02 | 0.05 | 0.02 |
| | $GLS$ | 0.05 | 0.00 | 0.03 | 0.02 | 0.06 | 0.02 |
| | $GMMpl^{(2)}$ | -0.06 | -0.01 | 0.05 | 0.02 | 0.08 | 0.03 |
| | $GMMpl^{(1)}$ | -0.02 | -0.00 | 0.06 | 0.03 | 0.06 | 0.03 |
| | $GMMld^{(2)}$ | 0.03 | 0.00 | 0.04 | 0.02 | 0.05 | 0.02 |
| | $GMMld^{(1)}$ | 0.02 | 0.00 | 0.06 | 0.03 | 0.07 | 0.03 |
| II | $LSDV$ | -0.14 | 0.00 | 0.04 | 0.02 | 0.14 | 0.02 |
| | $LSDVc$ | -0.02 | 0.00 | 0.04 | 0.02 | 0.05 | 0.02 |
| | $GLS$ | 0.07 | -0.01 | 0.03 | 0.02 | 0.08 | 0.03 |
| | $GMMpl^{(2)}$ | -0.08 | -0.01 | 0.05 | 0.03 | 0.10 | 0.03 |
| | $GMMpl^{(1)}$ | -0.02 | -0.00 | 0.06 | 0.03 | 0.07 | 0.03 |
| | $GMMld^{(2)}$ | 0.04 | -0.00 | 0.04 | 0.03 | 0.06 | 0.03 |
| | $GMMld^{(1)}$ | 0.03 | 0.00 | 0.06 | 0.03 | 0.07 | 0.04 |
| III | $LSDV$ | -0.17 | 0.01 | 0.04 | 0.03 | 0.17 | 0.03 |
| | $LSDVc$ | -0.04 | -0.02 | 0.04 | 0.03 | 0.06 | 0.04 |
| | $GLS$ | 0.11 | -0.05 | 0.03 | 0.03 | 0.12 | 0.06 |
| | $GMMpl^{(2)}$ | -0.11 | -0.00 | 0.06 | 0.03 | 0.12 | 0.03 |
| | $GMMpl^{(1)}$ | -0.03 | -0.00 | 0.07 | 0.03 | 0.08 | 0.03 |
| | $GMMld^{(2)}$ | 0.06 | -0.02 | 0.04 | 0.03 | 0.07 | 0.04 |
| | $GMMld^{(1)}$ | 0.04 | -0.01 | 0.06 | 0.04 | 0.07 | 0.04 |
| IV | $LSDV$ | -0.20 | -0.03 | 0.06 | 0.05 | 0.20 | 0.06 |
| | $LSDVc$ | 0.08 | 0.11 | 0.21 | 0.12 | 0.22 | 0.16 |
| | $GLS$ | 0.06 | -0.01 | 0.03 | 0.03 | 0.06 | 0.03 |
| | $GMMpl^{(2)}$ | -0.14 | -0.05 | 0.10 | 0.08 | 0.17 | 0.09 |
| | $GMMpl^{(1)}$ | -0.07 | -0.05 | 0.19 | 0.16 | 0.20 | 0.16 |
| | $GMMld^{(2)}$ | 0.04 | 0.01 | 0.04 | 0.04 | 0.06 | 0.04 |
| | $GMMld^{(1)}$ | 0.03 | 0.02 | 0.08 | 0.07 | 0.09 | 0.07 |
| V | $LSDV$ | -0.20 | 0.06 | 0.04 | 0.10 | 0.20 | 0.11 |
| | $LSDVc$ | -0.03 | 0.01 | 0.06 | 0.09 | 0.07 | 0.09 |
| | $GLS$ | 0.11 | -0.09 | 0.03 | 0.04 | 0.12 | 0.10 |
| | $GMMpl^{(2)}$ | -0.13 | 0.01 | 0.07 | 0.19 | 0.15 | 0.19 |
| | $GMMpl^{(1)}$ | -0.04 | -0.01 | 0.09 | 0.27 | 0.10 | 0.27 |
| | $GMMld^{(2)}$ | 0.05 | -0.03 | 0.04 | 0.08 | 0.07 | 0.09 |
| | $GMMld^{(1)}$ | 0.04 | -0.01 | 0.07 | 0.15 | 0.08 | 0.15 |
| VI | $LSDV$ | -0.21 | 0.05 | 0.04 | 0.05 | 0.21 | 0.07 |
| | $LSDVc$ | -0.04 | -0.07 | 0.06 | 0.06 | 0.07 | 0.09 |
| | $GLS$ | 0.14 | -0.13 | 0.03 | 0.04 | 0.14 | 0.14 |
| | $GMMpl^{(2)}$ | -0.14 | 0.01 | 0.07 | 0.06 | 0.16 | 0.06 |
| | $GMMpl^{(1)}$ | -0.05 | 0.01 | 0.10 | 0.07 | 0.11 | 0.07 |
| | $GMMld^{(2)}$ | 0.08 | -0.06 | 0.04 | 0.05 | 0.09 | 0.08 |
| | $GMMld^{(1)}$ | 0.05 | -0.03 | 0.07 | 0.06 | 0.09 | 0.07 |

Table 5: Simulation results for scheme 2, $T = 10$ and $N = 20$

| | | $Bias\ \gamma$ | $Bias\ \beta$ | $std\ \gamma$ | $std\beta$ | $RMSE\ \gamma$ | $RMSE\ \beta$ |
|---|---|---|---|---|---|---|---|
| VII | $LSDV$ | -0.10 | 0.01 | 0.05 | 0.03 | 0.12 | 0.03 |
| | $LSDVc$ | -0.00 | 0.01 | 0.06 | 0.03 | 0.06 | 0.03 |
| | $GLS$ | 0.07 | -0.01 | 0.04 | 0.03 | 0.08 | 0.03 |
| | $GMMpl^{(2)}$ | -0.11 | -0.00 | 0.07 | 0.03 | 0.13 | 0.03 |
| | $GMMpl^{(1)}$ | -0.03 | -0.00 | 0.09 | 0.04 | 0.09 | 0.04 |
| | $GMMld^{(2)}$ | 0.08 | -0.01 | 0.05 | 0.03 | 0.09 | 0.03 |
| | $GMMld^{(1)}$ | 0.05 | -0.00 | 0.08 | 0.04 | 0.09 | 0.04 |
| VIII | $LSDV$ | -0.14 | 0.00 | 0.06 | 0.04 | 0.15 | 0.04 |
| | $LSDVc$ | -0.03 | 0.00 | 0.07 | 0.04 | 0.07 | 0.04 |
| | $GLS$ | 0.07 | -0.01 | 0.04 | 0.03 | 0.08 | 0.04 |
| | $GMMpl^{(2)}$ | -0.14 | -0.01 | 0.08 | 0.04 | 0.16 | 0.04 |
| | $GMMpl^{(1)}$ | -0.05 | -0.01 | 0.10 | 0.05 | 0.11 | 0.05 |
| | $GMMld^{(2)}$ | 0.07 | -0.01 | 0.05 | 0.04 | 0.08 | 0.04 |
| | $GMMld^{(1)}$ | 0.05 | -0.00 | 0.08 | 0.05 | 0.09 | 0.05 |
| IX | $LSDV$ | -0.17 | -0.00 | 0.06 | 0.05 | 0.18 | 0.05 |
| | $LSDVc$ | -0.05 | -0.01 | 0.07 | 0.05 | 0.09 | 0.05 |
| | $GLS$ | 0.07 | -0.01 | 0.04 | 0.04 | 0.08 | 0.04 |
| | $GMMpl^{(2)}$ | -0.18 | -0.02 | 0.08 | 0.06 | 0.20 | 0.06 |
| | $GMMpl^{(1)}$ | -0.07 | -0.02 | 0.12 | 0.07 | 0.14 | 0.07 |
| | $GMMld^{(2)}$ | 0.06 | -0.00 | 0.05 | 0.05 | 0.08 | 0.05 |
| | $GMMld^{(1)}$ | 0.05 | -0.00 | 0.08 | 0.07 | 0.09 | 0.07 |
| X | $LSDV$ | -0.19 | 0.08 | 0.07 | 0.12 | 0.20 | 0.15 |
| | $LSDVc$ | -0.05 | 0.05 | 0.09 | 0.11 | 0.10 | 0.12 |
| | $GLS$ | 0.13 | -0.14 | 0.04 | 0.05 | 0.14 | 0.15 |
| | $GMMpl^{(2)}$ | -0.21 | 0.07 | 0.09 | 0.18 | 0.23 | 0.20 |
| | $GMMpl^{(1)}$ | -0.11 | 0.05 | 0.15 | 0.32 | 0.19 | 0.32 |
| | $GMMld^{(2)}$ | 0.11 | -0.12 | 0.05 | 0.08 | 0.12 | 0.14 |
| | $GMMld^{(1)}$ | 0.08 | -0.10 | 0.09 | 0.15 | 0.12 | 0.18 |
| XI | $LSDV$ | -0.21 | 0.06 | 0.07 | 0.14 | 0.22 | 0.16 |
| | $LSDVc$ | -0.05 | 0.02 | 0.09 | 0.13 | 0.10 | 0.13 |
| | $GLS$ | 0.11 | -0.09 | 0.05 | 0.07 | 0.12 | 0.11 |
| | $GMMpl^{(2)}$ | -0.22 | 0.03 | 0.10 | 0.22 | 0.24 | 0.22 |
| | $GMMpl^{(1)}$ | -0.10 | -0.01 | 0.15 | 0.40 | 0.18 | 0.40 |
| | $GMMld^{(2)}$ | 0.08 | -0.06 | 0.05 | 0.09 | 0.10 | 0.11 |
| | $GMMld^{(1)}$ | 0.06 | -0.04 | 0.09 | 0.18 | 0.11 | 0.19 |
| XII | $LSDV$ | -0.21 | 0.03 | 0.07 | 0.17 | 0.22 | 0.17 |
| | $LSDVc$ | -0.05 | -0.02 | 0.09 | 0.15 | 0.11 | 0.15 |
| | $GLS$ | 0.08 | -0.04 | 0.05 | 0.07 | 0.09 | 0.09 |
| | $GMMpl^{(2)}$ | -0.22 | -0.02 | 0.10 | 0.26 | 0.24 | 0.26 |
| | $GMMpl^{(1)}$ | -0.10 | -0.07 | 0.14 | 0.49 | 0.17 | 0.49 |
| | $GMMld^{(2)}$ | 0.06 | -0.02 | 0.06 | 0.10 | 0.08 | 0.10 |
| | $GMMld^{(1)}$ | 0.04 | -0.01 | 0.10 | 0.20 | 0.11 | 0.20 |

Table 6: Simulation results for scheme 2, $T = 10$ and $N = 50$

| | | $Bias\ \gamma$ | $Bias\ \beta$ | $std\ \gamma$ | $std\beta$ | $RMSE\ \gamma$ | $RMSE\ \beta$ |
|---|---|---|---|---|---|---|---|
| VII | $LSDV$ | -0.10 | 0.01 | 0.03 | 0.02 | 0.11 | 0.02 |
| | $LSDVc$ | -0.00 | 0.01 | 0.04 | 0.02 | 0.04 | 0.02 |
| | $GLS$ | 0.07 | -0.01 | 0.03 | 0.02 | 0.07 | 0.02 |
| | $GMMpl^{(2)}$ | -0.05 | -0.00 | 0.05 | 0.02 | 0.07 | 0.02 |
| | $GMMpl^{(1)}$ | -0.01 | -0.00 | 0.05 | 0.02 | 0.06 | 0.02 |
| | $GMMld^{(2)}$ | 0.04 | -0.00 | 0.04 | 0.02 | 0.06 | 0.02 |
| | $GMMld^{(1)}$ | 0.03 | 0.00 | 0.06 | 0.03 | 0.06 | 0.03 |
| VIII | $LSDV$ | -0.14 | 0.00 | 0.04 | 0.02 | 0.14 | 0.02 |
| | $LSDVc$ | -0.02 | 0.00 | 0.04 | 0.02 | 0.05 | 0.02 |
| | $GLS$ | 0.07 | -0.01 | 0.03 | 0.02 | 0.08 | 0.03 |
| | $GMMpl^{(2)}$ | -0.08 | -0.01 | 0.05 | 0.03 | 0.10 | 0.03 |
| | $GMMpl^{(1)}$ | -0.02 | -0.00 | 0.06 | 0.03 | 0.07 | 0.03 |
| | $GMMld^{(2)}$ | 0.04 | -0.00 | 0.04 | 0.03 | 0.06 | 0.03 |
| | $GMMld^{(1)}$ | 0.03 | 0.00 | 0.06 | 0.03 | 0.07 | 0.04 |
| IX | $LSDV$ | -0.17 | -0.00 | 0.04 | 0.03 | 0.17 | 0.03 |
| | $LSDVc$ | -0.04 | -0.01 | 0.05 | 0.03 | 0.06 | 0.03 |
| | $GLS$ | 0.08 | -0.01 | 0.02 | 0.03 | 0.08 | 0.03 |
| | $GMMpl^{(2)}$ | -0.11 | -0.02 | 0.06 | 0.04 | 0.13 | 0.04 |
| | $GMMpl^{(1)}$ | -0.03 | -0.01 | 0.07 | 0.04 | 0.08 | 0.04 |
| | $GMMld^{(2)}$ | 0.04 | -0.00 | 0.04 | 0.03 | 0.05 | 0.03 |
| | $GMMld^{(1)}$ | 0.03 | 0.00 | 0.06 | 0.05 | 0.07 | 0.05 |
| X | $LSDV$ | -0.19 | 0.08 | 0.04 | 0.08 | 0.19 | 0.11 |
| | $LSDVc$ | -0.03 | 0.05 | 0.06 | 0.07 | 0.07 | 0.09 |
| | $GLS$ | 0.14 | -0.15 | 0.03 | 0.03 | 0.14 | 0.15 |
| | $GMMpl^{(2)}$ | -0.13 | 0.05 | 0.07 | 0.15 | 0.15 | 0.16 |
| | $GMMpl^{(1)}$ | -0.05 | 0.02 | 0.09 | 0.20 | 0.10 | 0.20 |
| | $GMMld^{(2)}$ | 0.07 | -0.08 | 0.04 | 0.07 | 0.08 | 0.11 |
| | $GMMld^{(1)}$ | 0.05 | -0.06 | 0.07 | 0.13 | 0.08 | 0.14 |
| XI | $LSDV$ | -0.20 | 0.06 | 0.04 | 0.10 | 0.20 | 0.11 |
| | $LSDVc$ | -0.03 | 0.01 | 0.06 | 0.09 | 0.07 | 0.09 |
| | $GLS$ | 0.11 | -0.09 | 0.03 | 0.04 | 0.12 | 0.10 |
| | $GMMpl^{(2)}$ | -0.13 | 0.01 | 0.07 | 0.19 | 0.15 | 0.19 |
| | $GMMpl^{(1)}$ | -0.04 | -0.01 | 0.09 | 0.27 | 0.10 | 0.27 |
| | $GMMld^{(2)}$ | 0.05 | -0.03 | 0.04 | 0.08 | 0.07 | 0.09 |
| | $GMMld^{(1)}$ | 0.04 | -0.01 | 0.07 | 0.15 | 0.08 | 0.15 |
| XII | $LSDV$ | -0.21 | 0.03 | 0.04 | 0.11 | 0.21 | 0.11 |
| | $LSDVc$ | -0.04 | -0.03 | 0.06 | 0.10 | 0.07 | 0.10 |
| | $GLS$ | 0.09 | -0.05 | 0.03 | 0.05 | 0.09 | 0.07 |
| | $GMMpl^{(2)}$ | -0.13 | -0.05 | 0.07 | 0.23 | 0.15 | 0.23 |
| | $GMMpl^{(1)}$ | -0.04 | -0.05 | 0.09 | 0.34 | 0.10 | 0.35 |
| | $GMMld^{(2)}$ | 0.04 | 0.00 | 0.04 | 0.09 | 0.06 | 0.09 |
| | $GMMld^{(1)}$ | 0.03 | 0.01 | 0.07 | 0.16 | 0.08 | 0.16 |

Table 7: Simulation results varying $\mu$, $SNR$ and $T$, $N = 50$

| | | $Bias\ \gamma$ | $Bias\ \beta$ | $std\ \gamma$ | $std\beta$ | $RMSE\ \gamma$ | $RMSE\ \beta$ |
|---|---|---|---|---|---|---|---|
| XIII, $T = 10$ | $LSDV$ | -0.14 | 0.00 | 0.04 | 0.02 | 0.14 | 0.02 |
| | $LSDVc$ | -0.02 | 0.00 | 0.04 | 0.02 | 0.05 | 0.02 |
| | $GLS$ | 0.07 | -0.01 | 0.03 | 0.02 | 0.08 | 0.03 |
| | $GMMpl^{(2)}$ | -0.08 | -0.01 | 0.05 | 0.03 | 0.10 | 0.03 |
| | $GMMpl^{(1)}$ | -0.02 | -0.00 | 0.06 | 0.03 | 0.07 | 0.03 |
| | $GMMld^{(2)}$ | 0.04 | -0.00 | 0.04 | 0.03 | 0.06 | 0.03 |
| | $GMMld^{(1)}$ | 0.03 | 0.00 | 0.06 | 0.03 | 0.07 | 0.04 |
| XIV, $T = 10$ | $LSDV$ | -0.06 | 0.00 | 0.02 | 0.01 | 0.07 | 0.01 |
| | $LSDVc$ | -0.01 | 0.00 | 0.02 | 0.01 | 0.03 | 0.01 |
| | $GLS$ | 0.02 | -0.00 | 0.02 | 0.01 | 0.03 | 0.01 |
| | $GMMpl^{(2)}$ | -0.03 | -0.00 | 0.03 | 0.01 | 0.04 | 0.01 |
| | $GMMpl^{(1)}$ | -0.01 | -0.00 | 0.04 | 0.02 | 0.04 | 0.02 |
| | $GMMld^{(2)}$ | 0.02 | -0.00 | 0.03 | 0.01 | 0.03 | 0.01 |
| | $GMMld^{(1)}$ | 0.01 | 0.00 | 0.04 | 0.01 | 0.04 | 0.01 |
| XV, $T = 10$ | $LSDV$ | -0.14 | 0.00 | 0.04 | 0.02 | 0.14 | 0.02 |
| | $LSDVc$ | -0.03 | 0.00 | 0.05 | 0.02 | 0.06 | 0.02 |
| | $GLS$ | 0.24 | -0.05 | 0.01 | 0.02 | 0.24 | 0.06 |
| | $GMMpl^{(2)}$ | -0.12 | -0.01 | 0.07 | 0.03 | 0.13 | 0.03 |
| | $GMMpl^{(1)}$ | -0.10 | -0.02 | 0.14 | 0.04 | 0.17 | 0.04 |
| | $GMMld^{(2)}$ | 0.21 | -0.01 | 0.02 | 0.03 | 0.21 | 0.03 |
| | $GMMld^{(1)}$ | 0.19 | 0.01 | 0.04 | 0.04 | 0.19 | 0.04 |
| XVI, $T = 10$ | $LSDV$ | -0.06 | 0.00 | 0.02 | 0.01 | 0.07 | 0.01 |
| | $LSDVc$ | -0.01 | 0.00 | 0.03 | 0.01 | 0.03 | 0.01 |
| | $GLS$ | 0.20 | -0.02 | 0.01 | 0.01 | 0.20 | 0.03 |
| | $GMMpl^{(2)}$ | -0.04 | -0.00 | 0.04 | 0.01 | 0.05 | 0.01 |
| | $GMMpl^{(1)}$ | -0.03 | -0.00 | 0.07 | 0.02 | 0.07 | 0.02 |
| | $GMMld^{(2)}$ | 0.17 | -0.01 | 0.03 | 0.02 | 0.18 | 0.02 |
| | $GMMld^{(1)}$ | 0.14 | 0.01 | 0.04 | 0.02 | 0.15 | 0.02 |
| XIII, $T = 5$ | $LSDV$ | -0.29 | -0.01 | 0.06 | 0.04 | 0.30 | 0.04 |
| | $LSDVc$ | -0.08 | -0.00 | 0.10 | 0.04 | 0.13 | 0.04 |
| | $GLS$ | 0.08 | -0.01 | 0.03 | 0.03 | 0.08 | 0.03 |
| | $GMMpl^{(2)}$ | -0.12 | -0.03 | 0.13 | 0.07 | 0.18 | 0.07 |
| | $GMMpl^{(1)}$ | -0.06 | -0.02 | 0.15 | 0.07 | 0.16 | 0.08 |
| | $GMMld^{(2)}$ | 0.03 | -0.00 | 0.07 | 0.05 | 0.07 | 0.05 |
| | $GMMld^{(1)}$ | 0.02 | -0.00 | 0.09 | 0.06 | 0.10 | 0.06 |
| XIII, $T = 20$ | $LSDV$ | -0.06 | 0.01 | 0.02 | 0.02 | 0.07 | 0.02 |
| | $LSDVc$ | -0.00 | 0.00 | 0.02 | 0.01 | 0.02 | 0.01 |
| | $GLS$ | 0.06 | -0.01 | 0.02 | 0.01 | 0.06 | 0.02 |
| | $GMMpl^{(2)}$ | -0.05 | 0.00 | 0.03 | 0.02 | 0.06 | 0.02 |
| | $GMMpl^{(1)}$ | -0.01 | 0.00 | 0.03 | 0.02 | 0.03 | 0.02 |
| | $GMMld^{(2)}$ | 0.06 | -0.01 | 0.02 | 0.02 | 0.07 | 0.02 |
| | $GMMld^{(1)}$ | 0.03 | 0.00 | 0.04 | 0.02 | 0.05 | 0.02 |