



TI 2002-101/4

Tinbergen Institute Discussion Paper

The Effects of Dynamic Feedbacks on LS and MM Estimator Accuracy in Panel Data Models

Maurice J.G. Bun

Jan F. Kiviet

Faculty of Economics and Econometrics, University of Amsterdam, and Tinbergen Institute

Tinbergen Institute

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

Tinbergen Institute Amsterdam

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

Tinbergen Institute Rotterdam

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

Most TI discussion papers can be downloaded at
<http://www.tinbergen.nl>

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)

19 February, 2004

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 a range of (generalized) method of moments (MM) estimators in panel data models with individual effects and both a lagged dependent variable regressor and another explanatory variable. The latter may be affected by lagged feedbacks from the dependent variable too. Asymptotic expansions indicate how the order of magnitude of bias of MM estimators tends to increase with the number of moment conditions exploited. They also provide analytic evidence on how the bias of the various estimators depends on the feedbacks and on other model characteristics such as prominence of individual effects and correlation between observed and unobserved heterogeneity. Simulation results corroborate the theoretical findings and reveal that in small samples of models with dynamic feedbacks none of the techniques examined dominates regarding bias and mean squared error over all parametrizations examined.

1. Introduction

Economic relationships usually involve dynamic adjustment processes. In time-series regression models it is common practice to deal with 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

* Tinbergen Institute & Faculty of Economics and Econometrics, Universiteit van Amsterdam, Roetersstraat 11, 1018 WB Amsterdam, The Netherlands (+31 20 525 4217, M.J.G.Bun@UvA.NL, J.F.Kiviet@UvA.NL). We thank the editors and three anonymous referees for their constructive remarks. We are also grateful for helpful comments by various colleagues, Frank Windmeijer in particular. The first author wants to thank Cemmap (IFS, University College, London) and the second author the universities of Illinois (Champaign-Urbana), Montréal (CRDE), Hong Kong (HKUST), Melbourne and Sydney for hospitality while working on this paper.

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 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; 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 with respect to the current disturbance term. In this paper we will focus on the bias and efficiency of estimators for dynamic panel data models with individual effects and white-noise (not necessarily Gaussian) disturbances, which include a lagged dependent variable and another explanatory variable, and we distinguish the two cases where – conditional on the individual effects – this latter variable is either strongly exogenous or weakly exogenous due to lagged feedbacks. The particular exogeneity properties of both regressors imply a range of valid moment conditions and thus define 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 choosing between the various LS and MM estimators when the goal is to obtain accurate 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 moment conditions and employs the asymptotic covariance of these moment conditions as a weight matrix. It has been observed, though, that in moderately large samples consistent – especially asymptotically most efficient – estimators may show substantial bias, and that actual MSE (mean squared error) may deteriorate by using an abundance of moment conditions, see Ziliak (1997) and Koenker and Machado (1999).

In order to obtain a better understanding of these phenomena in the context of panel data we will apply in this study higher-order asymptotic methods to obtain for the various estimators the leading term in an expansion of their estimation error. This will uncover the asymptotic order with respect to both T and N of the bias of the various estimators and it will disclose also 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 term of its expansion will contain information on the parameters that prominently determine any bias in finite sample. In addition to these analytical investigations,

the actual finite sample bias, standard deviation and MSE of the various LS and MM estimators will be assessed in a series of Monte Carlo experiments. The theoretical findings will also prove to be helpful in designing the simulation experiments because they provide the major determining factors of the actual finite sample behaviour.

In Kiviet (1999) higher-order asymptotic methods have been applied to LSDV and to particular simple MM methods in dynamic panel data models with predetermined regressors which 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 a particular GMM estimator. 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 particular circumstances. However, clear-cut guidelines for practitioners are not yet available, especially not for the situation where N and T are both small and regressors next to the lagged dependent variable may be weakly and not strongly exogenous. Some Monte Carlo results for models with instantaneous feedbacks are presented in Blundell, Bond and Windmeijer (2000).

We shall produce simulation evidence for models with various forms of lagged feedbacks and with correlation between observed and unobserved heterogeneity and compare actual bias with a numerical evaluation of the higher-order asymptotic approximation of the bias. We do this for LSDV, GLS and for various forms of 1-step (G)MM estimators, including the so-called system estimator, and compare the numerical and the analytic effects of using all available instruments, amounting to a number of order $O(T^2)$, with omitting instruments and use only a number of order $O(T)$ of them, or just a fixed $O(1)$ number.

The structure of this paper is as follows. After the introduction of the model and its particular stochastic structure in Section 2, and the presentation of the various estimators in Section 3, we obtain in Section 4 the leading terms of the estimation errors of the various estimators. There we examine for all estimators the order of magnitude of bias in terms of N and T and the effects on estimator location of various forms of lagged feedbacks and the prominence of the unobserved individual effects. In Section 5 appropriate simulations are designed 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 6 concludes.

2. Stochastic structure of the model

We consider the standard first-order dynamic panel data model with only one further explanatory variable¹, i.e.

$$y_{it} = \gamma y_{i,t-1} + \beta x_{it} + u_{it}, \quad i = 1, \dots, N; t = 1, \dots, T, \quad (2.1)$$

where the disturbance term $u_{it} = \eta_i + \varepsilon_{it}$ contains two error components, viz. an unobserved individual specific effect η_i and a general disturbance term ε_{it} . We assume that the time-variant regressor x_{it} may be correlated with η_i and is predetermined with respect to ε_{it} , i.e.

$$\left. \begin{aligned} \mathbf{E}(x_{is}\varepsilon_{it}) &= 0, & s \leq t \\ \mathbf{E}(x_{is}\varepsilon_{it}) &\neq 0, & s > t \end{aligned} \right\} \quad i = 1, \dots, N. \quad (2.2)$$

Below we will formalize the correlation of x_{it} with η_i and the lagged feedback mechanism of past disturbances on x_{it} . We assume mutual independence of the cross-section units and serial independence of the disturbances, i.e.

$$\left. \begin{aligned} \eta_i &\sim i.i.d.(0, \sigma_\eta^2) \\ \varepsilon_{it} &\sim i.i.d.(0, \sigma_\varepsilon^2) \end{aligned} \right\} \quad i = 1, \dots, N; t = 0, \dots, T, \quad (2.3)$$

and hence we also suppose that, probably after a successful weighting operation, the disturbances are homoskedastic, both in the time and in the cross-section dimension. We define ε_{i0} because it will enable to specify the random characteristics of the start-up values y_{i0} and lagged feedbacks in x_{i1} . As usual we assume

$$\mathbf{E}(\eta_i \varepsilon_{jt}) = 0, \quad \forall i, j, t, \quad (2.4)$$

$$\mathbf{E}(y_{i0} \varepsilon_{jt}) = 0, \quad \forall i, j, t > 0, \quad (2.5)$$

hence the two error components are uncorrelated and 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. $|\gamma| < 1$.

In order to obtain higher-order asymptotic results we need expressions that make fully explicit how all observations on y_{it} and x_{it} depend on both error components. As in Kiviet (1999) we therefore 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 η_i and the disturbance terms ε_{it} . In the analysis to follow we shall condition on $\bar{x}_{it} = x_{it} - \tilde{x}_{it}$, $\bar{y}_{it} = y_{it} - \tilde{y}_{it}$ and on $\bar{y}_{i0} = y_{i0} - \tilde{y}_{i0}$, $\forall i, t$.

Regarding x we start by examining the same simple setup as in Kiviet (1999) where x_{it} can be decomposed as

$$\left. \begin{aligned} x_{it} &= \bar{x}_{it} + \tilde{x}_{it} \\ \tilde{x}_{it} &= \phi \varepsilon_{i,t-1} + \pi \eta_i \end{aligned} \right\} \quad i = 1, \dots, N; t = 1, \dots, T, \quad (2.6)$$

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

¹Our findings can be generalized – though will not yield qualitatively different results as it seems – for models with higher-order lags and more explanatories. For ease of exposition we thus restrict ourselves to the relatively simple model (2.1).

exogenous, and when $\phi \neq 0$ then x_{it} is weakly exogenous, because, conditional on η_i and the past, the joint density of y_{it} and x_{it} can be factorized in the density of y_{it} conditional on x_{it} and the marginal density of x_{it} such that the parameters of these two densities are unrelated. The form of weak exogeneity in (2.6) is very specific, but it captures the essentials of lagged feedbacks, whereas its simplicity makes further derivations relatively easy. Later we will discuss and simulate more general forms of weak exogeneity.

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

$$\left. \begin{aligned} \tilde{y}_{it} &= \gamma \tilde{y}_{i,t-1} + \beta \tilde{x}_{it} + \eta_i + \varepsilon_{it} \\ \bar{y}_{it} &= \gamma \bar{y}_{i,t-1} + \beta \bar{x}_{it} \end{aligned} \right\} \quad i = 1, \dots, N; \quad t = 1, \dots, T. \quad (2.7)$$

To be able to decompose the relevant random components of \tilde{y}_{it} into the two error components η_i and ε_{it} we have to make an assumption on the accumulated size of the individual effect η_i in y_{i0} . For the sake of simplicity we shall assume that the full long-run impact of the individual effect η_i on y_{it} is already present in y_{i0} , which implies

$$\mathbf{E}(\tilde{y}_{it} \mid \eta_i) = \alpha \eta_i, \quad i = 1, \dots, N; \quad t = 0, \dots, T, \quad (2.8)$$

where $\alpha = \frac{1+\beta\pi}{1-\gamma}$, so we assume effect-stationarity (labelled mean-stationarity in Blundell and Bond, 1998) for both x_{it} and y_{it} . For the N initial values we further assume

$$\tilde{y}_{i0} = \alpha \eta_i + \omega \varepsilon_{i0}, \quad \forall i \quad (2.9)$$

where ω is either 0 or 1. In case of lagged feedback (i.e. $\phi \neq 0$) x_{i1} depends on ε_{i0} and then \tilde{y}_{i0} should also contain ε_{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 strongly 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 ε_{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 ($i = 1, \dots, N$)

$$y_i = \gamma y_{i(-1)} + \beta x_i + \eta_i \iota_T + \varepsilon_i, \quad (2.10)$$

$$x_i = \bar{x}_i + \phi \varepsilon_{i(-1)} + \pi \eta_i \iota_T, \quad (2.11)$$

where $y_{i(-1)} = (y_{i0}, \dots, y_{i,T-1})'$, $\varepsilon_{i(-1)} = (\varepsilon_{i0}, \dots, \varepsilon_{i,T-1})'$ and $\iota_T = (1, \dots, 1)'$ 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 $\Gamma_T = (I_T - \gamma L_T)^{-1}$ and using (2.9), the relevant random part of y_i is

$$\tilde{y}_i = \alpha \eta_i \iota_T + \Gamma_T (I_T + \beta \phi L_T) \varepsilon_i + (\omega \gamma + \phi \beta) \Gamma_T e_{T,1} \varepsilon_{i0}. \quad (2.12)$$

Stacking the T observations per individual over all N individuals yields

$$y = W \delta + u, \quad (2.13)$$

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 $\eta = (\eta_1, \dots, \eta_N)'$. Using (2.6) and (2.12) the relevant random parts of y and x can be written explicitly as

$$\tilde{y} = \alpha S\eta + \Gamma(I_{NT} + \beta\phi L)\varepsilon + (\omega\gamma + \phi\beta)\Gamma(I_N \otimes e_{T,1})\varepsilon_0, \quad (2.14)$$

$$\tilde{x} = \pi S\eta + \phi[L\varepsilon + (I_N \otimes e_{T,1})\varepsilon_0], \quad (2.15)$$

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

The expressions (2.14) and (2.15) will be used for obtaining analytical results on the properties of the various estimators of δ 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

$$J_T = \begin{pmatrix} 0 & 1 & 0 & \cdot & \cdot & 0 \\ 0 & 0 & 1 & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & \cdot & \cdot & \cdot & 0 & 1 \end{pmatrix}, \quad K_T = \begin{pmatrix} 1 & 0 & \cdot & \cdot & \cdot & 0 \\ 0 & 1 & 0 & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & 1 & 0 & \cdot & \cdot \\ 0 & \cdot & \cdot & 0 & 1 & 0 \end{pmatrix} \quad (2.16)$$

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 δ and η in (2.13) by OLS yields estimates which are called Least Squares Dummy Variables (LSDV), fixed effect or within group estimates. For δ this estimator can be expressed as

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

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. This estimator can also be obtained by premultiplying model (2.13) by any matrix $R = I_N \otimes R_T$, where R_T is $(T-1) \times T$ with rank $T-1$ and $R_T\iota_T = 0$. While removing the individual effects such a transformation reduces the number of observations and produces disturbances with $\mathbf{V}(R\varepsilon) = \sigma_\varepsilon^2 RR'$, which is full rank. Applying GLS yields $[W'R'(RR')^{-1}RW]^{-1}W'R'(RR')^{-1}Ry = \hat{\delta}_{\text{LSDV}}$, because $R'(RR')^{-1}R = A$ follows from the orthonormality of $[T^{-1/2}\iota_T, R_T'(R_T R_T')^{-1/2}]$. Valid matrices R_T are D_T (first differences) and P_T (forward orthogonal deviations), see Arellano and Bover (1995). The latter transformation will prove to be useful below when constructing and analyzing MM estimators. The $(T-1) \times T$ upper-triangular matrix P_T transforms as

$$y_{it}^* = c_t[y_{it} - (T-t)^{-1}(y_{i,t+1} + \dots + y_{iT})], \quad (3.2)$$

with $c_t^2 = (T-t)/(T-t+1)$. Since $P_T P_T' = I_{T-1}$, independence of ε_{it} is preserved in the transformed model.

Treating the individual effects as random, the covariance matrix of the combined disturbance term $u = S\eta + \varepsilon$ is $V = \sigma_\eta^2 S S' + \sigma_\varepsilon^2 I_{NT}$ and the Generalized Least Squares

(GLS) estimator of δ is

$$\begin{aligned}\hat{\delta}_{\text{GLS}} &= (W'V^{-1}W)^{-1}W'V^{-1}y, \text{ where} \\ V^{-1} &= I_N \otimes (I_T - \theta \frac{1}{T} \iota_T \iota_T'), \quad \theta = 1 - \left(1 + T \frac{\sigma_\eta^2}{\sigma_\varepsilon^2}\right)^{-1}.\end{aligned}\tag{3.3}$$

Note that $0 \leq \theta \leq 1$, and $\theta \rightarrow 1$ as $T \rightarrow \infty$ for $\sigma_\eta^2 > 0$, which shows the equivalence of GLS and LSDV for T large. However, for finite T these estimators differ and V^{-1} will only partially wipe out the individual effects, which explains that the GLS finite sample bias is not invariant with respect to $\sigma_\eta^2/\sigma_\varepsilon^2$, whereas LSDV's is under effect-stationarity. Note, moreover, that GLS is not feasible unless $\sigma_\eta^2/\sigma_\varepsilon^2$ is known. In our simulations we will also consider the feasible FGLS estimator

$$\hat{\delta}_{\text{FGLS}} = (W\hat{V}^{-1}W)^{-1}W'\hat{V}^{-1}y,\tag{3.4}$$

where \hat{V}^{-1} is estimated by obtaining $\hat{\theta}$ from estimates $\hat{\sigma}_\varepsilon^2$ and $\hat{\sigma}_\eta^2$ based on within and between residuals as in Doornik et al. (2002).

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, see Ahn and Schmidt (1995). In this study we will focus on MM implementations using linear moment conditions only and we will not exploit any moment conditions associated with the homoskedasticity of ε_{it} . Arellano and Bond (1991) used all linear moment conditions for the model in first differences associated with the predeterminedness of the regressors and the absence of serial correlation in the (original) disturbances. Blundell and Bond (1998) suggest to exploit in addition moment conditions associated with the effect-stationarity of the variables. The various sets of linear moment conditions emerge straightforwardly, depending on whether or how the individual effects are removed from the model or from the instruments.

3.2.1. Removing the effects from the model

Like Anderson and Hsiao (1982), Arellano and Bond (1991) suggest to take first differences, i.e.

$$Dy = DW\delta + D\varepsilon.\tag{3.5}$$

Since $\mathbf{E}[(Dy_{(-1)})'D\varepsilon] = -\mathbf{E}(y'_{(-1)}\varepsilon_{(-1)}) = -\mathbf{E}(\varepsilon'_{(-1)}\varepsilon_{(-1)}) = -\sigma_\varepsilon^2 N(T-1)$, the correlation of one of the regressors with the errors of (3.5) is such that here LS is inconsistent, irrespective of how the sample size is extended. However, there are many valid moment conditions for individual i in the differenced model (3.5), viz.

$$\left. \begin{aligned}\mathbf{E}(y_{i,t-s}\Delta\varepsilon_{it}) &= 0, & t = 2, \dots, T; & \quad s = 2, \dots, t, \\ \mathbf{E}(x_{i,t-s}\Delta\varepsilon_{it}) &= 0, & t = 2, \dots, T; & \quad s = 1, \dots, t-1.\end{aligned}\right\}\tag{3.6}$$

Even more valid instruments than these $T(T-1)$ are available under our particular form of weak exogeneity (2.6), or when x is strongly exogenous, but we will not use these here. Not just D , but any transformation matrix R as defined above, for example P , will eliminate the effects from the model, see Arellano and Honoré (2001, p.3255). For model

$$Ry = RW\delta + R\varepsilon,\tag{3.7}$$

we shall consider GMM estimation exploiting $m \leq T(T - 1)$ moment conditions for each individual i . These can be expressed as $\mathbf{E}(Z'_i R_T \varepsilon_i) = 0$, where Z_i is a $(T - 1) \times m$ matrix with variables in levels. We will consider three particular instrument matrices Z_i , viz. the block-diagonal matrix (with increasing block size) $Z_i^{(2)}$, which includes all $m^{(2)} = T(T - 1) = O(T^2)$ instruments given in (3.6), i.e.

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

the block-diagonal matrix (with constant block size) $Z_i^{(1)}$, which includes only a subset of $m^{(1)} = 2(T - 1) = O(T)$ of the $Z_i^{(2)}$ instruments, viz.

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

and the matrix $Z_i^{(0)}$, which includes a linear transformation of the instruments in $Z_i^{(2)}$, reducing their number further to $m^{(0)} = 4 = O(T^0)$, viz.

$$Z_i^{(0)} = \begin{pmatrix} y_{i0} & x_{i1} & 0 & 0 \\ y_{i1} & x_{i2} & y_{i0} & x_{i1} \\ \vdots & \vdots & \vdots & \vdots \\ y_{i,T-2} & x_{i,T-1} & y_{i,T-3} & x_{i,T-2} \end{pmatrix}.$$

Stacking over individuals one has for any m -columns instruments-in-levels matrix Z_l the conditions $\mathbf{E}(Z'_l R \varepsilon) = 0$, where $Z_l = (Z'_{l1}, \dots, Z'_{lN})'$ is $N(T - 1) \times m$. Provided $R' Z_l$ has full column rank, the optimal one-step GMM estimator is then

$$\hat{\delta}_{\text{GMM}} = [W' R' Z_l (Z'_l R R' Z_l)^{-1} Z'_l R W]^{-1} W' R' Z_l (Z'_l R R' Z_l)^{-1} Z'_l R y, \quad (3.8)$$

where $(Z'_l R R' Z_l)^{-1}$ provides a consistent (up to a scalar) estimator of the inverse of $\mathbf{E}(Z'_l R \varepsilon \varepsilon' R' Z_l)$ for $\varepsilon_{it} \sim i.i.d.$ Hence, when all available moment conditions (3.6) are exploited, i.e. $Z_l = Z_l^{(2)}$, estimator (3.8) is asymptotically efficient. For $R = P$ and writing $y^* = P y$ and $W^* = P W$, estimator (3.8) specializes in

$$\hat{\delta}_{\text{GMMfl}} = [W'^* Z_l (Z'_l Z_l)^{-1} Z'_l W^*]^{-1} W'^* Z_l (Z'_l Z_l)^{-1} Z'_l y^*, \quad (3.9)$$

where the acronym GMMfl makes clear that the model underwent the forward orthogonal deviations transformation (f) and uses level instruments (l). For $Z_l = Z_l^{(2)}$ this estimator is equivalent to the estimator of Arellano and Bond (1991) which uses $R = D$, see Arellano (2003, p.153). Since form (3.9) has proved to be more suitable for further analytical examination, see Alvarez and Arellano (2003) who analyzed it for $Z_l = Z_l^{(2)}$ in the panel AR(1) model, we will focus here on transformation P too. We will investigate it for

$Z_i = Z_i^{(h)}$ ($h = 0, 1, 2$), indicating the resulting estimators by $\hat{\delta}_{\text{GMMfd}^{(h)}}$, thus clarifying which instrument matrix $Z_i^{(h)}$ (h conforms to the power of T in the order of magnitude of the number of instruments) has been used. From the special structure of $Z_i^{(2)}$ it follows that a necessary condition for the existence of $\hat{\delta}_{\text{GMMfd}^{(2)}}$ is that $N \geq K(T - 1)$, where K is the number of columns of W , i.e. in our special case $T \leq 1 + N/2$.

3.2.2. Removing the effects from the instruments

For the equation in levels (2.13) first differences of current and lagged explanatory variables are valid instruments, provided that the variables are effect-stationary, see Arellano and Bover (1995), Kiviet (1995) and Blundell and Bond (1998). For individual i we then have the moment conditions

$$\left. \begin{aligned} \mathbf{E}(u_{it}\Delta y_{i,t-s}) &= 0, & t = 2, \dots, T; & \quad s = 1, \dots, t-1, \\ \mathbf{E}(u_{it}\Delta x_{i,t-s}) &= 0, & t = 2, \dots, T; & \quad s = 0, \dots, t-2. \end{aligned} \right\} \quad (3.10)$$

These, or linear transformations or reductions of them, can be expressed as $\mathbf{E}(Z'_{di}J_T u_i) = 0$, where the matrix Z_{di} is $(T - 1) \times m$. We will examine the three cases

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

which includes all $m^{(2)} = T(T - 1) = O(T^2)$ instruments-in-differences; its subset

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

where $m^{(1)} = 2(T - 1) = O(T)$; and the instrument matrix with $m^{(0)} = 4 = O(1)$

$$Z_{di}^{(0)} = \begin{pmatrix} \Delta y_{i1} & \Delta x_{i2} & 0 & 0 \\ \Delta y_{i2} & \Delta x_{i3} & \Delta y_{i1} & \Delta x_{i2} \\ \vdots & \vdots & \vdots & \vdots \\ \Delta y_{i,T-1} & \Delta x_{i,T} & \Delta y_{i,T-2} & \Delta x_{i,T-1} \end{pmatrix}.$$

Stacking over individuals one has $\mathbf{E}(Z'_d J u) = 0$, where $Z_d = (Z'_{d1}, \dots, Z'_{dN})'$ is either $Z_d^{(2)}$, $Z_d^{(1)}$ or $Z_d^{(0)}$. Here an optimal one-step GMM estimator would require a weight matrix involving the unobservable ratio $\sigma_\eta^2/\sigma_\varepsilon^2$. We will analyze here the particular operational one-step GMM estimator

$$\hat{\delta}_{\text{GMMld}} = [W' J' Z_d (Z'_d Z_d)^{-1} Z'_d J W]^{-1} W' J' Z_d (Z'_d Z_d)^{-1} Z'_d J y, \quad (3.11)$$

which has been considered too by Blundell, Bond and Windmeijer (2000). It is optimal only when $\sigma_\eta^2 = 0$. For $\hat{\delta}_{\text{GMMld}^{(h)}}$ ($h = 0, 1, 2$), where "ld" clarifies that the equation in levels has been instrumented by differenced variables, it is required that $(Z'_d Z_d)^{-1}$ exists. For $\hat{\delta}_{\text{GMMld}^{(2)}}$ this implicates $N \geq K(T - 1)$.

3.2.3. System GMM estimator

Combining in a system all the above moment conditions while omitting the redundant conditions, see Blundell and Bond (1998), yields

$$\mathbf{E}(Z'_{si}q_i) = 0, \quad \text{where } Z_{si} = \begin{pmatrix} Z_{li}^{(2)} & O \\ O & Z_{di}^{(1)} \end{pmatrix}, \quad q_i = \begin{pmatrix} P_T \varepsilon_i \\ J_T u_i \end{pmatrix}, \quad (3.12)$$

with Z_{si} a $2(T-1) \times m$ matrix and $m = (T+2)(T-1) = O(T^2)$. Defining

$$Z_s = \begin{pmatrix} Z_l^{(2)} & O \\ O & Z_d^{(1)} \end{pmatrix}, \quad W_s = \begin{pmatrix} PW \\ JW \end{pmatrix}, \quad y_s = \begin{pmatrix} Py \\ Jy \end{pmatrix}, \quad q_s = \begin{pmatrix} P\varepsilon \\ Ju \end{pmatrix}, \quad (3.13)$$

and using, like Blundell and Bond (1998), $Z'_s Z_s$ for weighting, we obtain

$$\hat{\delta}_{\text{GMMs}} = [W'_s Z_s (Z'_s Z_s)^{-1} Z'_s W_s]^{-1} W'_s Z_s (Z'_s Z_s)^{-1} Z'_s y_s. \quad (3.14)$$

This system GMM estimator requires $N \geq K(T-1)$.

4. Finite sample bias

We will now examine the location² of the various LS and MM estimators of δ by approximating their finite sample bias using asymptotic expansion techniques. For all these estimators the estimation error can be written as

$$\hat{\delta} - \delta = Q^{-1} F' v, \quad (4.1)$$

where Q is a 2×2 matrix, F is $n \times 2$, v is $n \times 1$, and n is either NT , $N(T-1)$ or $2N(T-1)$. Assuming that either or both N and T can get large and that all variables are stationary through time³, we have $Q = O_p(n)$, $\bar{Q} = \mathbf{E}(Q) = O(n)$ and $Q - \bar{Q} = O_p(n^{1/2})$, which yields $(Q - \bar{Q})\bar{Q}^{-1} = O_p(n^{-1/2})$. Expanding $Q^{-1} = \bar{Q}^{-1}[I + (Q - \bar{Q})\bar{Q}^{-1}]^{-1}$, we obtain

$$\begin{aligned} Q^{-1} &= \bar{Q}^{-1} - \bar{Q}^{-1}(Q - \bar{Q})\bar{Q}^{-1} + O_p(n^{-2}) \\ &= \bar{Q}^{-1} + O_p(n^{-3/2}), \end{aligned} \quad (4.2)$$

and, assuming $\mathbf{E}(F'v) = O(n^*)$, we find the first-order bias approximation

$$\mathbf{E}(\hat{\delta} - \delta) = \bar{Q}^{-1} \mathbf{E}(F'v) + O(n^* n^{-3/2}), \quad (4.3)$$

where $B = [\mathbf{E}(Q)]^{-1} \mathbf{E}(F'v) = O(n^* n^{-1})$ is the leading term of the bias of $\hat{\delta}$. Whether $n = NT$, $N(T-1)$ or $2N(T-1)$ has no practical consequences here, so any differences in the order of the leading term of the bias result from $\mathbf{E}(F'v)$, i.e. from n^* .

In Appendix A we show for LSDV that $\mathbf{E}(F'v) = \mathbf{E}(W'A\varepsilon) = O(NT^0)$, hence the bias is $O(N^0 T^{-1})$, indicating that LSDV is inconsistent for large N and finite T . More precisely, see also Bun and Kiviet (2003), $\mathbf{E}(\hat{\delta}_{\text{LSDV}} - \delta) = B_{\text{LSDV}} + O(N^{-1} T^{-1})$, with

$$\begin{aligned} B_{\text{LSDV}} &= \sigma_\varepsilon^2 \bar{Q}_{\text{LSDV}}^{-1} \{[\text{tr}(\Pi) + \beta\phi \text{tr}(\Pi L)]e_{2,1} + \phi \text{tr}(AL)e_{2,2}\} \\ &= -\sigma_\varepsilon^2 N \bar{Q}_{\text{LSDV}}^{-1} \left(\frac{1 + \beta\phi}{1 - \gamma} e_{2,1} + \phi e_{2,2} \right) + O(T^{-2}) = O(T^{-1}), \end{aligned} \quad (4.4)$$

²Since we employ always at least two more instruments than regressors, we may assume that the first two moments exist of all MM estimators considered here.

³The analysis to follow can be generalized for non-stationary variables as in Kiviet (1995), but this would yield no new insights, provided that all variables remain effect-stationary with respect to the individual effects.

where $\bar{Q}_{\text{LSDV}} = \mathbf{E}(W'AW)$, $\Pi = AL\Gamma$, and in the final expression we substituted

$$\left. \begin{aligned} \text{tr}(\Pi) &= -\frac{N}{1-\gamma} \left(1 - \frac{1}{T} \frac{1-\gamma^T}{1-\gamma}\right), \\ \text{tr}(\Pi L) &= -\frac{N}{1-\gamma} \left[1 - \frac{1}{T} \left(1 + \frac{1-\gamma^{T-1}}{1-\gamma}\right)\right], \\ \text{tr}(AL) &= -N \left(1 - \frac{1}{T}\right). \end{aligned} \right\} \quad (4.5)$$

Note that B_{LSDV} can be obtained without assuming normality of the disturbances, because for evaluating $\mathbf{E}(W'A\varepsilon)$ and $\mathbf{E}(W'AW)$ the first two moments of ε suffice. Result (4.4) highlights that the $O(T^{-1})$ bias of LSDV is directly affected by both feedback parameters γ and ϕ , which affect \bar{Q}_{LSDV} too, whereas the LSDV bias is invariant with respect to $\sigma_\eta^2/\sigma_\varepsilon^2$ and π under effect-stationarity.

Along similar lines we find $\mathbf{E}(\hat{\delta}_{\text{GLS}} - \delta) = B_{\text{GLS}} + O(n^{-1})$, where

$$\begin{aligned} B_{\text{GLS}} &= \sigma_\varepsilon^2 \theta \bar{Q}_{\text{GLS}}^{-1} \left\{ \left[\text{tr}(\Pi) + \beta \phi \text{tr}(\Pi L) + N \frac{1 + \beta \pi}{1 - \gamma} \right] e_{2,1} - N \left[\phi \left(1 - \frac{1}{T}\right) - \pi \right] e_{2,2} \right\} \\ &= -\sigma_\varepsilon^2 \theta (\phi - \pi) N \bar{Q}_{\text{GLS}}^{-1} \left(\frac{\beta}{1 - \gamma} e_{2,1} + e_{2,2} \right) + O(T^{-2}) = O(T^{-1}), \end{aligned} \quad (4.6)$$

with $\bar{Q}_{\text{GLS}} = \mathbf{E}(W'V^{-1}W)$. Hence, the bias of GLS depends not only on γ , β , ϕ and σ_ε^2 , but, unlike LSDV, also on π and θ . The leading bias term disappears for $\theta \rightarrow 0$, which is not the case for LSDV. Otherwise ($\sigma_\eta^2 > 0$) the bias of GLS is affected by the relative magnitude of the error components $\sigma_\eta^2/\sigma_\varepsilon^2$ and by π , and therefore LSDV seems more robust than GLS. When $\phi = \pi$ the bias is $O(N^0 T^{-2})$; we conjecture that this does not hold in general, but results from the specific form of feedback (2.6).

Using instruments in levels we find $\mathbf{E}(\hat{\delta}_{\text{GMMff}^{(2)}} - \delta) = B_{\text{GMMff}^{(2)}} + O(n^{-1})$, with

$$\begin{aligned} B_{\text{GMMff}^{(2)}} &= 2\sigma_\varepsilon^2 \bar{Q}_{\text{GMMff}^{(2)}}^{-1} \sum_{s=2}^T \{ \text{tr}[A_s L_s \Gamma_s + \beta \phi A_s L_s \Gamma_s L_s] e_{2,1} + \phi \text{tr}(A_s L_s) e_{2,2} \} \\ &= -2T \sigma_\varepsilon^2 \bar{Q}_{\text{GMMff}^{(2)}}^{-1} \left(\frac{1 + \beta \phi}{1 - \gamma} e_{2,1} + \phi e_{2,2} \right) + O(N^{-1} T^{-1}) = O(N^{-1}). \end{aligned} \quad (4.7)$$

Here A_s , L_s and Γ_s are $s \times s$ matrices of similar structure as A_T , L_T and Γ_T . That $B_{\text{GMMff}^{(2)}} = O(N^{-1})$ in the panel AR(1) model was already found by Alvarez and Arellano (2003). From (4.7) we see that both sources of feedback in our more general model entail bias of order $O(N^{-1})$, also when $\beta = 0$. Reducing now the number of moment conditions to $O(T)$ we find $\mathbf{E}(\hat{\delta}_{\text{GMMff}^{(1)}} - \delta) = B_{\text{GMMff}^{(1)}} + O(N^{-1} T^{-2})$, with

$$\begin{aligned} B_{\text{GMMff}^{(1)}} &= 2\sigma_\varepsilon^2 \bar{Q}_{\text{GMMff}^{(1)}}^{-1} N^{-1} \{ [\text{tr}(\Pi) + \beta \phi \text{tr}(\Pi L)] e_{2,1} + \phi \text{tr}(AL) e_{2,2} \} \\ &= -2\sigma_\varepsilon^2 \bar{Q}_{\text{GMMff}^{(1)}}^{-1} \left(\frac{1 + \beta \phi}{1 - \gamma} e_{2,1} + \phi e_{2,2} \right) + O(N^{-1} T^{-2}) = O(N^{-1} T^{-1}). \end{aligned} \quad (4.8)$$

Hence, reducing the number of instruments by a factor T has reduced the leading term of the bias by a factor T too. This could make the actual bias considerably smaller if $\bar{Q}_{\text{GMMff}^{(2)}}^{-1}$ and $\bar{Q}_{\text{GMMff}^{(1)}}^{-1}$ do not differ very much. Also note that the factors in parenthesis in the simplified expressions of (4.4), (4.7) and (4.8) are equivalent and, apart from the difference in \bar{Q}^{-1} matrices, these leading terms differ by factors N , $2T$ and 2 respectively. Hence, for both N and T substantial $\text{GMMff}^{(1)}$ seems less vulnerable regarding bias.

Deriving the leading term of the bias of $\text{GMMfl}^{(0)}$ is very complex, but in Appendix A we prove that $\mathbf{E}(\hat{\delta}_{\text{GMMfl}^{(0)}} - \delta) = B_{\text{GMMfl}^{(0)}} + O(N^{-1}T^{-2})$, with

$$B_{\text{GMMfl}^{(0)}} = O(N^{-1}T^{-1}). \quad (4.9)$$

Hence, reducing the number of instruments further (and inhibit the block-diagonal structure of the Z_{li} matrices) no longer reduces the order of the GMMfl bias. From the proofs it is apparent that the determining factor for the order of bias is not the actual choice we made with respect to the reduction in number of instruments (we did remove the instruments with the longest lags), but that it is just the order of magnitude of their total number which determines the order of the bias.

Regarding GMMld we obtain $\mathbf{E}(\hat{\delta}_{\text{GMMld}^{(2)}} - \delta) = B_{\text{GMMld}^{(2)}} + O(N^{-1})$, with

$$B_{\text{GMMld}^{(2)}} = \sigma_{\eta}^2(T^2 + T - 2)\bar{Q}_{\text{GMMld}^{(2)}}^{-1} \left(\frac{1 + \beta\pi}{1 - \gamma} e_{2,1} + \pi e_{2,2} \right) = O(TN^{-1}), \quad (4.10)$$

which, surprisingly⁴, suggests a bias that increases with T . For $\mathbf{E}(\hat{\delta}_{\text{GMMld}^{(1)}} - \delta) = B_{\text{GMMld}^{(1)}} + O(N^{-1}T^{-1})$ we find

$$B_{\text{GMMld}^{(1)}} = 2\sigma_{\eta}^2(T - 1)\bar{Q}_{\text{GMMld}^{(1)}}^{-1} \left(\frac{1 + \beta\pi}{1 - \gamma} e_{2,1} + \pi e_{2,2} \right) = O(N^{-1}), \quad (4.11)$$

hence, again the bias has reduced by a factor T through reducing the number of instruments by a factor T . Note that the π in the factor in parenthesis for the GMMld expressions is a ϕ in the GMMfl expressions. Using just a constant number of instruments we find $\mathbf{E}(\hat{\delta}_{\text{GMMld}^{(0)}} - \delta) = B_{\text{GMMld}^{(0)}} + O(N^{-1}T^{-2})$ with

$$B_{\text{GMMld}^{(0)}} = O(N^{-1}T^{-1}). \quad (4.12)$$

Hence, here an even further bias reduction does prove possible by limiting the number of differenced instruments to a fixed number.

Finally, for GMMs we find $\mathbf{E}(\hat{\delta}_{\text{GMMs}} - \delta) = B_{\text{GMMs}} + O(N^{-1}T^{-1})$, with

$$\begin{aligned} B_{\text{GMMs}} &= \bar{Q}_{\text{GMMs}}^{-1} (\bar{Q}_{\text{GMMfl}^{(2)}} B_{\text{GMMfl}^{(2)}} + \bar{Q}_{\text{GMMld}^{(1)}} B_{\text{GMMld}^{(1)}}) \\ &= -2\sigma_{\varepsilon}^2 T \bar{Q}_{\text{GMMs}}^{-1} \left[\frac{1 - \frac{\sigma_{\eta}^2}{\sigma_{\varepsilon}^2}}{1 - \gamma} e_{2,1} + \left(\phi - \pi \frac{\sigma_{\eta}^2}{\sigma_{\varepsilon}^2} \right) \left(\frac{\beta}{1 - \gamma} e_{2,1} + e_{2,2} \right) \right] = O(N^{-1}). \end{aligned} \quad (4.13)$$

Hence, GMMs, which is a matrix weighted average of $\text{GMMfl}^{(2)}$ and $\text{GMMld}^{(1)}$, has a bias of similar order as these two estimators. However, in the panel AR(1) model, where the second term between the square brackets in (4.13) is void, the leading term of the bias of GMMs is in fact just of order $O(N^{-1}T^{-1})$, provided $\sigma_{\eta}^2 = \sigma_{\varepsilon}^2$. As it happens, this is the case in simulation designs in both Blundell and Bond (1998) and Doornik et al. (2002), which gave rise to the conclusion that the performance of GMMs is superior to $\text{GMMfl}^{(2)}$. In the simulations below we will examine their relative performance when $\sigma_{\eta}^2 \neq \sigma_{\varepsilon}^2$ and when the model contains extra regressors which may be correlated in some way with the error components.

⁴The difference in order of magnitude of finite sample bias of the GMMfl and GMMld estimators could be the result of the fact that the latter estimator does not exploit an optimal weight matrix. From the study of the panel AR(1) model in Alvarez and Arellano (2003) it follows that using a non-optimal weight matrix in the Arellano-Bond estimator $\text{GMMfl}^{(2)}$ increases the order of its bias to $O(TN^{-1})$ too.

The results on the order of finite sample bias in models with weakly exogenous regressors obtained here have been summarized in Table 1⁵. Note that the LSDV estimator can be obtained from all three model transformations distinguished. Despite their equivalence in order of bias, GLS is found to be less robust than LSDV as its estimation error depends on both $\sigma_\eta^2/\sigma_\varepsilon^2$ and π . The same has been established for the GMMld estimators and for GMMs. Note that for $h = 1, 2$ the order of bias of $\text{GMMfl}^{(h)}$ is a factor T smaller than for $\text{GMMld}^{(h)}$, whereas it is equivalent for $\text{GMMfl}^{(1)}$, $\text{GMMfl}^{(0)}$ and $\text{GMMld}^{(0)}$. The latter three have all smaller order of bias than GMMs. Hence, since $\sigma_\eta^2/\sigma_\varepsilon^2$ and π are generally unknown, we conclude from these higher-order asymptotic results that in samples with both T and N moderate or large the $\text{GMMld}^{(0)}$, $\text{GMMfl}^{(1)}$ and $\text{GMMfl}^{(0)}$ estimators seem preferable over all others considered as far as bias is concerned. When T is large relative to N the $\text{GMMfl}^{(2)}$, $\text{GMMld}^{(2)}$ and GMMs estimators do not exist and LSDV, GLS and $\text{GMMfl}^{(h)}$ ($h \leq 1$) seem preferable, and when T is small relative to N the GMMfl methods seem preferable to LSDV and GLS. Especially $\text{GMMld}^{(2)}$ does not seem to be recommendable 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 GMMs, the other GMMld and all GMMfl variants, because these have all a bias of the same order in N . The leading terms of the bias seem to indicate that the bias increases with γ for all estimators. Note, however, that the bias approximations have been obtained assuming $|\gamma| < 1$, implying, for instance, $\gamma^T = o(T^0)$ and hence they do not allow to infer what happens for $\gamma \rightarrow 1$. The effects of γ on the size of the bias seem of smaller order for GLS when $\beta = 0$ or $\phi = 0$ and $\pi = 0$, and also for GMMs when $\sigma_\eta^2 = \sigma_\varepsilon^2$ and $\phi = 0$, $\pi = 0$. Finally note that GMMfl is not fully invariant regarding $\sigma_\eta^2/\sigma_\varepsilon^2$ and π , as LSDV is, since these parameters affect $Z_l(Z_l'Z_l)^{-1}Z_l'$.

5. Simulation results

Data for y have been generated according to equation (2.1) with two different models for the weakly exogenous explanatory variable x . In scheme 1 its generating equation is designed as in (2.6) with \bar{x}_{it} a stationary AR(1) process, i.e.

$$\left. \begin{aligned} x_{it}^{(1)} &= \bar{x}_{it} + \phi_1 \varepsilon_{i,t-1} + \pi_1 \eta_i \\ \bar{x}_{it} &= \rho_1 \bar{x}_{i,t-1} + \xi_{it} \end{aligned} \right\} \quad i = 1, \dots, N; \quad t = 1, \dots, T, \quad (5.1)$$

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

$$x_{it}^{(1)} = \frac{1}{1 - \rho_1 \mathcal{L}} \xi_{it} + \phi_1 \varepsilon_{i,t-1} + \pi_1 \eta_i, \quad (5.2)$$

where \mathcal{L} is the lag operator. For $|\rho_1| < 1$ this process is stationary. When $\phi_1 = 0$ regressor $x^{(1)}$ is in fact strongly exogenous.

In the earlier sections scheme 1 proved to be relatively easy to handle from an analytical point of view. However, from an empirical point of view a more interesting alternative

⁵We have not omitted the factors N^0 and T^0 in the order expressions to stress that we examined asymptotic behavior for both $N \rightarrow \infty$ and $T \rightarrow \infty$. Note that $O(N^{-1})$ is in agreement with $O(N^{-1}T^\alpha)$ for any real α when keeping T fixed.

weakly exogenous process, indicated as scheme 2, is

$$x_{it}^{(2)} = \rho_2 x_{i,t-1}^{(2)} + \phi_2 y_{i,t-1}^{(2)} + \pi_2 \eta_i + \xi_{it}, \quad (5.3)$$

which is again strongly exogenous when $\phi_2 = 0$. When $\phi_2 \neq 0$, $x_{it}^{(2)}$ depends via $y_{i,t-1}^{(2)}$ on all past disturbances and not just on $\varepsilon_{i,t-1}$. Scheme 2 implies

$$x_{it}^{(2)} = \frac{(1 - \gamma \mathcal{L})\xi_{it} + \phi_2 \varepsilon_{i,t-1}}{1 - (\gamma + \beta \phi_2 + \rho_2)\mathcal{L} + \gamma \rho_2 \mathcal{L}^2} + \frac{(1 - \gamma)\pi_2 + \phi_2}{(1 - \gamma)(1 - \rho_2) - \beta \phi_2} \eta_i, \quad (5.4)$$

which, like (5.2), is a linear combination of three stochastically independent components, but now an ARMA(2,1) process based on ξ_{it} , an AR(2) process based on $\varepsilon_{i,t-1}$ and the individual effects. However, in scheme 2 these three components are each determined by all parameters, which is not the case for scheme 1. In (5.4) not just the value of ϕ_2 itself, but also the values of γ , β and ρ_2 affect the extent of the feedback mechanism in x when $\phi_2 \neq 0$. Therefore, when γ is varied in the simulation, not only the speed of adjustment, but also the characteristics of x_{it} (its variance, correlogram, feedback pattern and magnitude of individual effect) will vary, which complicates the interpretation of the simulation results. For stationarity of $x_{it}^{(2)}$ it is required that the three restrictions

$$\left. \begin{array}{l} \gamma \rho_2 < 1 \\ \gamma + \rho_2(1 - \gamma) + \beta \phi_2 < 1 \\ \gamma + \rho_2(1 + \gamma) + \beta \phi_2 > -1 \end{array} \right\} \quad (5.5)$$

hold jointly, which they do as is shown below.

We chose $\gamma = \{0.25, 0.75\}$ and the long-run effect $\beta/(1 - \gamma)$ of x on y has been set equal to unity in all experiments, implying $\beta = 1 - \gamma$. Further, we chose $\rho_1, \rho_2 = \{0.5, 0.95\}$, which yield stationary regressors. 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 $\phi_1 = \phi_2/(1 - \gamma - \beta \phi_2 - \rho_2 + \gamma \rho_2)$ implying

$$\phi_2 = \frac{\phi_1(1 - \gamma)(1 - \rho_2)}{1 + \beta \phi_1}. \quad (5.6)$$

Substituting $\beta = 1 - \gamma$ and (5.6) in the second restriction of (5.5) yields $\gamma + (\rho_2 + \phi_2)(1 - \gamma) < 1$, which implies that the parameters should obey $\rho_2 + \phi_2 < 1$, which means

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

and this is true. Hence, $x_{it}^{(2)}$ is stationary, because our parameter choices also obey the other two restrictions of (5.5). We also took $\pi_1 = \{-1, 0, +1\}$ and to advance comparability we therefore took $\pi_1 = [(1 - \gamma)\pi_2 + \phi_2]/[(1 - \gamma)(1 - \rho_2) - \beta \phi_2]$ implying

$$\pi_2 = \pi_1(1 - \rho_2 - \phi_2) - \frac{\phi_2}{1 - \gamma}. \quad (5.7)$$

We shall normalize with respect to the variance of the disturbance term σ_ε^2 . Hence, the design parameters that remain to be fixed now are σ_η^2 and σ_ξ^2 . Appendix B gives further details on how we attributed values to these two variances. We chose a relationship

between σ_η^2 and σ_ε^2 such that the impact on $\mathbf{V}(y_{it})$ of the two variance components η_i and ε_{it} has a ratio μ^2 . This implies for scheme 1

$$\sigma_\eta^2 = \mu^2 \frac{(1-\gamma)(1+2\gamma\beta\phi_1 + \beta^2\phi_1^2)}{(1+\gamma)(1+\beta\pi_1)^2} \quad (5.8)$$

and for scheme 2

$$\begin{aligned} \sigma_\eta^2 = \mu^2 & \left(1 + \rho_2^2 - 2\rho_2 \frac{\gamma + \beta\phi_2 + \rho_2}{1 + \gamma\rho_2} \right) \left[\frac{1 - \rho_2 + \beta\pi_2}{(1-\gamma)(1-\rho_2) - \beta\phi_2} \right]^{-2} \times \\ & \left[1 - \gamma^2\rho_2^2 - \frac{1 - \gamma\rho_2}{1 + \gamma\rho_2} (\gamma + \beta\phi_2 + \rho_2)^2 \right]^{-1}. \end{aligned} \quad (5.9)$$

In the simulations we examined $\mu = \{0, 1, 5\}$. The parameter σ_ξ^2 has been determined by controlling the signal-to-noise ratio (ζ) 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 $s = \{1, 2\}$

$$\zeta^{(s)} = \frac{\mathbf{V}(y_{it}^{(s)} - \varepsilon_{it} \mid \eta_i)}{\mathbf{V}(\varepsilon_{it})} = \frac{\mathbf{V}(y_{it}^{(s)} \mid \eta_i)}{\mathbf{V}(\varepsilon_{it})} - 1. \quad (5.10)$$

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

$$\sigma_\xi^2 = \frac{1}{\beta^2} \left[\zeta^{(1)} - \frac{(\gamma + \beta\phi_1)^2}{1 - \gamma^2} \right] \frac{(1 - \gamma^2)(1 - \rho_1^2)(1 - \gamma\rho_1)}{1 + \gamma\rho_1} \quad (5.11)$$

and scheme 2

$$\begin{aligned} \sigma_\xi^2 = \frac{1}{\beta^2} & (\zeta^{(2)} + 1) \left[1 - \gamma^2\rho_2^2 - \frac{1 - \gamma\rho_2}{1 + \gamma\rho_2} (\gamma + \beta\phi_2 + \rho_2)^2 \right] \\ & - \frac{1}{\beta^2} \left(1 + \rho_2^2 - 2\rho_2 \frac{\gamma + \beta\phi_2 + \rho_2}{1 + \gamma\rho_2} \right). \end{aligned} \quad (5.12)$$

Selfevidently, not every combination of values of the parameters is feasible.

In all experiments the individual effects η_i and the (other) stochastic source for the x_{it} variables ξ_{it} were drawn from the normal distribution. For the noise $\varepsilon_{it} \sim i.i.d.(0, 1)$, however, besides the standard normal distribution, we also chose one with fatter tails, viz. $t_3/\sqrt{3}$ (standardized Student with 3 degrees of freedom) and a skew one, viz. $(\chi_1^2 - 1)/\sqrt{2}$. Regarding sample size we will focus on rather small values, viz. $T = \{5, 10, 20, 50\}$ and $N = \{20, 50\}$. The range of values assigned above to the design parameters, i.e. to γ , β , ρ_s , ϕ_s , π_s , σ_η^2 (through μ), σ_ξ^2 (through ζ) with $\sigma_\varepsilon^2 = 1$, define 216 different experiments. These have not all been examined for both scheme $s = \{1, 2\}$, different distributions of ε_{it} , and sample sizes N, T . For some selected parametrizations the Tables 2 through 6 present Monte Carlo estimates (each based on 10,000 replications) for the true bias, standard error (*ster*) and root mean squared error (*rmse*) of the various estimators of γ and β . In the tables for scheme 1 \bar{B}_γ and \bar{B}_β give assessments of the bias obtained from the leading terms of our expansions, where the matrices \bar{Q} have been obtained as an average over the simulations of the relevant moment matrices, whereas true parameter values have been

substituted in our approximations for $E(F'v)$. Comparison of \bar{B}_γ and \bar{B}_β with the actual (estimated) bias reveals the accuracy of the higher-order approximations.

Tables 2 and 3 both concern the model according to scheme 1 with normal disturbances at sample size $T = 10$, $N = 20$ for $\gamma = 0.75$. In Table 2 the x_{it} variable is strongly exogenous and not correlated with the individual effects. In designs 1 through 4 it has variance such that the signal-to-noise ratio ζ is 3 and in designs 1 and 2 the relative magnitude of the individual effects is such that $\mu = 1$, which here implies $\sigma_\eta^2 = \frac{1}{7}\sigma_\varepsilon^2$, see (5.8). The only difference between designs 1 and 2 is the value of ρ . We see that biases and rmse's are larger when x_{it} is smoother. Both designs show that the bias of LSDV and of $\text{GMMf}^{(2)}$ is substantial and comparable in sign and magnitude, while $\text{GMMf}^{(1)}$ is less biased, which is in agreement with our theoretical findings. The larger variance of $\text{GMMf}^{(1)}$ in comparison to $\text{GMMf}^{(2)}$ is in agreement with the latter's asymptotic efficiency. However, applying a rmse criterion and focussing on γ the $\text{GMMf}^{(1)}$ estimator is more efficient than LSDV and $\text{GMMf}^{(2)}$, and LSDV is more efficient than $\text{GMMf}^{(2)}$, especially in design 2. We note that $\text{GMMf}^{(0)}$ has a much smaller bias than the other estimators that use the forward orthogonal deviation transformation but at the expense of larger variance. The GLS and GMMld estimators, which all start off from the levels equation, tend to show a bias of opposite sign, and the non-operational GLS estimator performs very well here, though FGLS is worse than $\text{GMMld}^{(2)}$. The latter shows its efficiency superiority with respect to the variants exploiting fewer instruments, though it is more severely biased, which is again in agreement with our higher-order asymptotic findings. As expected, the GMMs estimator is not showing an exceptionally small bias (although it seems to benefit from the opposite sign in bias of its building blocks $\text{GMMf}^{(2)}$ and $\text{GMMld}^{(1)}$), but regarding rmse it comes close to $\text{GMMld}^{(2)}$. In all further results reported we stick to the $\rho = 0.95$ case. In designs 3 and 4 we just examine the effects of changing μ . In design 3 there are no individual effects and then the estimators that use the levels equation perform very well with respect to bias, which is in agreement with their derived leading bias term being zero. The other estimators, including GMMs, do not benefit from $\sigma_\eta^2 = 0$ and they show a large bias, especially regarding γ , and a poor rmse, especially for β . In design 4 we have $\mu = 5$, i.e. $\sigma_\eta^2 = 3\frac{4}{7}\sigma_\varepsilon^2$. GLS can cope with that, whereas all other estimators, apart from those using a fixed number of instruments, are heavily biased. Of the operational methods $\text{GMMld}^{(2)}$ has smallest rmse for $\mu = 1$, but at $\mu = 5$ even LSDV performs better. In designs 5 and 6 we examine the effects of increasing the signal-to-noise ratio. We find that the biases are smaller now, but they are still substantial for most estimators. From Table 2 we also see (as well as from all other tables below on scheme 1) that the leading term of the bias (which we did not derive for FGLS, $\text{GMMf}^{(0)}$ and $\text{GMMld}^{(0)}$) is usually reasonably accurate for the actual bias, especially for the LS estimators, despite the fact that the sample size is not very large. Note that we calculated these \bar{B} values from the initial crude expressions for the leading term of the bias, and not from the simplified form. Otherwise, the bias approximations for GLS would have been zero when $\phi = \pi$, which would have made them less precise, as we see.

In Table 3 we have $\gamma = 0.75$, $\zeta = 3$, $\rho = 0.95$ and we examine the effects of varying the exogeneity status of the regressor x_{it} . In designs 7 and 8 the regressor is not correlated with the effects, but it is no longer strongly exogenous. Upon comparing with design 2 we find that the bias and efficiency of estimating β has improved through the changes in the correlogram of x_{it} ; the bias of γ does not change very much as a rule. In designs 9 and 10 the regressor is again strongly exogenous but now we introduce correlation between

regressor and individual effects. Selfevidently we establish that LSDV is invariant with respect to π . The other estimators are affected by π but not all in the same way. That especially the bias of GMMld⁽²⁾ and GMMld⁽¹⁾ varies with π is again in agreement with the characteristics of their leading bias term. In designs 11 and 12 both π and ϕ are non-zero. We see that the numerical effects of both π and ϕ on bias are very moderate. Hence, we find that despite the equivalence in the order of bias due to the presence of either a lagged dependent variable or to indirect feedbacks via a weakly exogenous regressor, the actual magnitude of the bias seems to stem primarily from the direct feedbacks when $\gamma = 0.75$. Note that, like in Table 2, GMMs is performing relatively well, although it has a substantial bias, especially for μ away from one.

In Table 4 we look at $\gamma = 0.25$. Comparing designs 13 with 2, and 14 with 4, we see that especially FGLS and GMMld⁽²⁾ are much worse now. Although GMMs seems the best choice for practitioners in design 13, in design 14 we see that for $\mu = 5$ and γ small GMMs is much worse than the Arellano-Bond estimator. Especially the GMMld estimators are affected by ϕ and π . Note that over these designs GMMfl⁽⁰⁾ has always a small bias, whereas this is not the case for the β estimator of GMMld⁽⁰⁾ when $\pi = -1$.

Next we did rerun the designs 2, 4 and 11 upon varying respectively the distribution of the disturbances and the sample sizes T and N . For fat tailed and skew disturbances almost similar results as those for the normal were found, so to save space we do not report them. Increasing N we found that the bias of LSDV and GLS remains more or less the same, while the bias in the GMM estimators decreases. 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. For $N = 50$ we examined various values of T and report the results for design 11 in Table 5. We find indeed that increasing T aggravates the bias of GMMld⁽²⁾, has little effect on GMMld⁽¹⁾ and reduces the bias for all other techniques, as predicted by the leading term of the bias. Note that the estimators using $O(T^2)$ instruments do not exist for $N = T = 50$. Although in Tables 2 through 5 the bias approximations are not always very accurate, especially not for some MM estimators, there seem to be options for improving the estimators by performing some form of bias correction.

Finally in Table 6 we present some results for the more realistic scheme 2, which only differs from scheme 1 when $\phi \neq 0$. We note that for particular techniques the standard error and bias of the estimator for β changes substantially, which must be due to the change in correlogram of x_{it} . However, it seems that the simple scheme 1 is capable of representing the major phenomena that are also at stake when more general forms of feedbacks are present as incorporated in scheme 2.

6. Concluding remarks

By asymptotic expansion techniques we have produced important new analytical insights into the bias of the major least squares and method of moments estimators for panel data models with both lagged dependent variable regressors and additional strongly or weakly exogenous (predetermined) explanatory variables in the presence of unobserved individual effects and white-noise disturbances. In addition to establishing that LSDV and GLS are biased to order $O(T^{-1})$, irrespective of the value of N , whereas all the MM estimators are biased to order $O(N^{-1})$, assuming T fixed, we revealed more subtle differences between the various MM implementations. We examined two variants of MM estimators using either the levels equation or the model in forward orthogonal deviations, denoted by

GMMld and GMMfl respectively, and also the hybrid system estimator GMMs. GMMld and GMMfl have both been examined for the situation where all linear moment conditions associated with the predeterminedness of the regressors are being used (which amounts to a number of instruments of order T^2) and when fewer instruments are being exploited, viz. a number of order T , or just a fixed number. We found: (i) that both GMMld and GMMfl when employing all available linear moment conditions are biased to an order larger in magnitude by a factor T than when using a set of instruments which contain a number of instruments reduced to order T ; however, (ii) the aggravation of bias when the number of instruments is increased from order T to order T^2 is not affecting GMMld and GMMfl in the same way, because the bias in GMMfl is overall of smaller order in T than for GMMld; (iii) when a fixed number of instruments is used both GMMfl and GMMld have the same order of bias as GMMfl with T instruments; and (iv) this bias is of lower order than for the GMM system estimator GMMs, although we established that the latter can have smaller order of bias too under very specific parametric restrictions; (v) for all estimators we found that the bias due to feedbacks in the explanatory variable (weak instead of strong exogeneity) is in principle of similar order as the bias due to the presence of a lagged dependent variable regressor; (vi) our theoretical derivations also indicate that for (F)GLS, GMMld and GMMs the leading term of the bias is strongly affected by the magnitude of the individual effects and (vii) as well by any correlation of the regressor and the effects, which seems a serious drawback in practice.

We designed appropriate Monte Carlo experiments to assess whether all these qualitative differences are of practical numerical relevance in actual finite samples. In general, the simulations corroborate the theoretical findings. 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 programmed in a Monte Carlo study, would be very difficult to analyze analytically, and for the same reasons it leads to interpretation difficulties in a simulation. However, from a few experiments we infer that there seems no reason to suspect that there are principle differences between the results for these two schemes. The major results from the Monte Carlo study are as follows: (i) all findings seem little affected by deviations of the error distribution from the normal; (ii) they show for all MM estimators a reduction in bias and standard error when N is increased from 20 to 50, while the bias for LSDV and (F)GLS remains virtually unchanged; (iii) increasing T decreases the bias of LSDV, GLS, GMMfl and GMMld⁽⁰⁾, it has little effect on GMMld⁽¹⁾, GMMs and FGLS and indeed it aggravates the bias of GMMld⁽²⁾; (iv) increasing the number of moment conditions used in estimation increases the bias in finite samples considerably; (v) when γ is substantial, the magnitude of the bias changes moderately when the other regressor is weakly instead of strongly exogenous; (vi) apart from LSDV all estimators are found to have a bias determined by correlation of observed and unobserved heterogeneity; (vii) the performance of the FGLS and GMMld estimators depends heavily on the magnitude of the individual effects, occasionally leading to dramatic biases. The latter nuisance also affects the hybrid GMMs estimator.

Since for particular parameter values all techniques show substantial bias and poor rmse in samples where both T and N are moderate or small, standard first-order asymptotic theory is of little use indeed to establish and rank the qualities of estimators. The higher-order asymptotic findings of this study prove to be much more informative about the actual finite sample behaviour of the various methods. Especially FGLS and GMMld are not invariant with respect to μ and π , which are unknown in practice. A straightfor-

ward advise for practitioners regarding which method to prefer in small samples does not emerge, but under effect-stationarity GMMs seems a relatively safe choice, except when γ is small while the effects are prominent. In future research we will examine for this class of models for which parameter values and sample sizes our bias approximations – or approximations obtained via alternative asymptotic sequences – can be used effectively for bias correction and ultimately lead to more accurate inference.

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., 2003. The time series and cross-section asymptotics of dynamic panel data estimators. *Econometrica* 71, 1121-1159.
- Anderson, T.W., Hsiao, C., 1982. Formulation and Estimation of Dynamic Models using Panel Data. *Journal of Econometrics* 18, 47-82.
- Arellano, M., 2003. *Panel Data Econometrics*. Oxford University Press.
- 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.
- Bun, M.J.G., Kiviet, J.F., 2003. On the diminishing returns of higher-order terms in asymptotic expansions of bias. *Economics Letters*, 79, 145-152.
- Doornik, J.A., Arellano, M., Bond, S. (2002). Panel data estimation using DPD for Ox. mimeo, University of Oxford.
- 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. Derivation of the bias approximations

From (2.14) it follows that

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

where $\tilde{y}_0 = (\tilde{y}_{10}, \dots, \tilde{y}_{N0})'$ has been defined in (2.9). With (2.15) this gives

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

and also the relevant stochastic elements of any instrument matrices.

Using these, we find regarding LSDV

$$\mathbf{E}(W'A\varepsilon) = \mathbf{E}(\tilde{W}'A\varepsilon) + \mathbf{E}(\bar{W}'A\varepsilon) = \mathbf{E}(\tilde{W}'A\varepsilon),$$

because $\mathbf{E}(\bar{W}'A\varepsilon) = 0$. Using $\Pi = AL\Gamma$, we find

$$\begin{aligned}\mathbf{E}(W'A\varepsilon) &= \mathbf{E}[\varepsilon'\Pi(I_{NT} + \beta\phi L)\varepsilon]e_{2,1} + \phi\mathbf{E}(\varepsilon'AL\varepsilon)e_{2,2} \\ &= \sigma_\varepsilon^2 \text{tr}(\Pi)e_{2,1} + \sigma_\varepsilon^2\beta\phi \text{tr}(\Pi L)e_{2,1} + \sigma_\varepsilon^2\phi \text{tr}(AL)e_{2,2} = O(N),\end{aligned}\tag{A.3}$$

because ε_0 and ε are independent by assumption. This gives (4.4).

Regarding GLS we obtain

$$\mathbf{E}(W'V^{-1}u) = \mathbf{E}(\tilde{W}'V^{-1}u) + \mathbf{E}(\bar{W}'V^{-1}u) = \mathbf{E}(\tilde{W}'V^{-1}S\eta) + \mathbf{E}(\tilde{W}'V^{-1}\varepsilon).$$

Employing (A.2) we find

$$\begin{aligned}\mathbf{E}(\tilde{W}'V^{-1}S\eta) &= \mathbf{E}(\eta'S'V^{-1}S\eta)(\alpha e_{2,1} + \pi e_{2,2}) = \sigma_\eta^2 \text{tr}(S'V^{-1}S)(\alpha e_{2,1} + \pi e_{2,2}) \\ &= \sigma_\varepsilon^2 N\theta(\alpha e_{2,1} + \pi e_{2,2}),\end{aligned}$$

and in a similar way as in (A.3) and substituting $V^{-1} = I_N \otimes (I_T - \theta\frac{1}{T}\iota_T\iota_T')$ we obtain

$$\begin{aligned}\mathbf{E}(\tilde{W}'V^{-1}\varepsilon) &= \sigma_\varepsilon^2 \text{tr}(V^{-1}L\Gamma)e_{2,1} + \sigma_\varepsilon^2\beta\phi \text{tr}(V^{-1}L\Gamma L)e_{2,1} + \sigma_\varepsilon^2\phi \text{tr}(V^{-1}L)e_{2,2} \\ &= \sigma_\varepsilon^2\theta[\text{tr}(\Pi) + \beta\phi \text{tr}(\Pi L)]e_{2,1} - \sigma_\varepsilon^2\phi\theta N\left(1 - \frac{1}{T}\right)e_{2,2} = O(N),\end{aligned}$$

and these lead to (4.6).

Next, we examine the bias of $\hat{\delta}_{\text{GMMff}(h)}$. For $h = \{1, 2\}$ it is possible to rewrite (3.9) in an alternative convenient form. Let G 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, because the $Z_{li}^{(1)}$ and $Z_{li}^{(2)}$ are, GZ_l is a block-diagonal matrix, with blocks $Z'_{l1}, \dots, Z'_{l,T-1}$, where Z_{lt} is $N \times m_{lt}^{(h)}$ with $m_{lt}^{(2)} = 2t$ and $m_{lt}^{(1)} = 2$. Of course, G is orthonormal, i.e. $G'G = I_{N(T-1)}$ and $G^{-1} = G'$, hence

$$(Z'_l Z_l)^{-1} = (Z'_l G' G Z_l)^{-1} = \text{diag}[(Z'_{l1} Z_{l1})^{-1}, \dots, (Z'_{l,T-1} Z_{l,T-1})^{-1}],$$

so that one can write, see also Arellano and Bover (1995), for $h = 1, 2$

$$\hat{\delta}_{\text{GMM}}^{(h)} = \left[\sum_{t=1}^{T-1} W_t^{*'} Z_{lt} (Z'_{lt} Z_{lt})^{-1} Z'_{lt} W_t^* \right]^{-1} \sum_{t=1}^{T-1} W_t^{*'} Z_{lt} (Z'_{lt} Z_{lt})^{-1} Z'_{lt} y_t^*,$$

where $y_t^* = (y_{1t}^*, \dots, y_{Nt}^*)'$ and $W_t^* = (w_{1t}^*, \dots, w_{Nt}^*)'$. Now we derive for $h > 0$

$$\mathbf{E}(W_t^{*'} P' M_{Z_l} P \varepsilon) = \sum_{t=1}^{T-1} \mathbf{E}(W_t^{*'} M_{Z_{lt}} \varepsilon_t^*),$$

where $M_{Z_{lt}} = Z_{lt} (Z'_{lt} Z_{lt})^{-1} Z'_{lt}$ with $\text{tr}(M_{Z_{lt}}) = m_{lt}^{(h)}$. We proceed as follows, see also Alvarez and Arellano (2003). We have

$$\begin{aligned} \mathbf{E}(W_t^{*'} M_{Z_{lt}} \varepsilon_t^*) &= \mathbf{E} \begin{pmatrix} y_{t-1}' M_{Z_{lt}} \varepsilon_t^* \\ x_t' M_{Z_{lt}} \varepsilon_t^* \end{pmatrix} = \mathbf{E} \begin{pmatrix} \text{tr}(y_{t-1}' M_{Z_{lt}} \varepsilon_t^*) \\ \text{tr}(x_t' M_{Z_{lt}} \varepsilon_t^*) \end{pmatrix} \\ &= \mathbf{E} \mathbf{E}_{t-1} \begin{pmatrix} \text{tr}(M_{Z_{lt}} \varepsilon_t^* y_{t-1}') \\ \text{tr}(M_{Z_{lt}} \varepsilon_t^* x_t') \end{pmatrix}, \end{aligned}$$

where \mathbf{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

$$\mathbf{E} \mathbf{E}_{t-1} \begin{pmatrix} \text{tr}(M_{Z_{lt}} \varepsilon_t^* y_{t-1}') \\ \text{tr}(M_{Z_{lt}} \varepsilon_t^* x_t') \end{pmatrix} = \begin{pmatrix} \mathbf{E}\{\text{tr}[M_{Z_{lt}} \mathbf{E}_{t-1}(\varepsilon_t^* y_{t-1}')] \} \\ \mathbf{E}\{\text{tr}[M_{Z_{lt}} \mathbf{E}_{t-1}(\varepsilon_t^* x_t')] \} \end{pmatrix}.$$

Since ε_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. for any i

$$\begin{aligned} \mathbf{E}_{t-1}(\varepsilon_t^* y_{t-1}') &= \mathbf{E}(\varepsilon_t^* y_{t-1}') = I_N \mathbf{E}(\varepsilon_{it}^* y_{i,t-1}') \\ \mathbf{E}_{t-1}(\varepsilon_t^* x_t') &= \mathbf{E}(\varepsilon_t^* x_t') = I_N \mathbf{E}(\varepsilon_{it}^* x_{it}'), \end{aligned}$$

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

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

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

$$\sum_{t=1}^{T-1} \mathbf{E}(W_t^{*'} M_{Z_{lt}} \varepsilon_t^*) = \begin{pmatrix} \text{tr}[\mathbf{E}(\varepsilon_i y_{i(-1)}) \sum_{t=1}^{T-1} m_{lt}^{(r)} p_t p_t'] \\ \text{tr}[\mathbf{E}(\varepsilon_i x_i') \sum_{t=1}^{T-1} m_{lt}^{(r)} p_t p_t'] \end{pmatrix}. \quad (\text{A.4})$$

Using (2.12) and (2.11) we obtain

$$\begin{aligned}\mathbf{E}(\varepsilon_i y'_{i(-1)}) &= \mathbf{E}(\varepsilon_i \tilde{y}'_{i(-1)}) = \mathbf{E}[\varepsilon_i \varepsilon'_i (I_T + \beta \phi L'_T) \Gamma'_T L'_T] = \sigma_\varepsilon^2 (I_T + \beta \phi L'_T) \Gamma'_T L'_T \\ \mathbf{E}(\varepsilon_i x'_i) &= \mathbf{E}(\varepsilon_i \tilde{x}'_i) = \phi \mathbf{E}(\varepsilon_i \varepsilon'_i L'_T) = \phi \sigma_\varepsilon^2 L'_T,\end{aligned}$$

and for $h = 2$ we have $\sum_{t=1}^{T-1} m_{it}^{(2)} p_t p'_t = 2 \sum_{t=1}^{T-1} t p_t p'_t = 2 \sum_{s=2}^T H_s A_s H'_s$, where $H_s = (O : I_s)'$ is a $T \times s$ selection matrix and $A_s = I_s - \frac{1}{s} \iota_s \iota'_s$ (see again Alvarez and Arellano, 2003), whereas for $h = 1$ we have $\sum_{t=1}^{T-1} m_{it}^{(1)} p_t p'_t = 2 \sum_{t=1}^{T-1} p_t p'_t = 2A_T$. For the first element of (A.4) we now find for $h = 2$

$$\begin{aligned}\text{tr}[\mathbf{E}(\varepsilon_i y'_{i(-1)}) \sum_{t=1}^{T-1} m_{it}^{(2)} p_t p'_t] &= 2\sigma_\varepsilon^2 \sum_{s=2}^T \text{tr}[(I_T + \beta \phi L'_T) \Gamma'_T L'_T H_s A_s H'_s] \\ &= 2\sigma_\varepsilon^2 \sum_{s=2}^T \text{tr}(A_s L_s \Gamma_s + \beta \phi A_s L_s \Gamma_s L_s) \\ &= -\frac{2}{1-\gamma} \sigma_\varepsilon^2 \sum_{s=2}^T \left\{ 1 - \frac{1}{s} \frac{1-\gamma^s}{1-\gamma} + \beta \phi \left[1 - \frac{1}{s} \left(1 + \frac{1-\gamma^{s-1}}{1-\gamma} \right) \right] \right\} \\ &= -\frac{2T(1+\beta\phi)}{1-\gamma} \sigma_\varepsilon^2 + o(1) = O(T),\end{aligned}$$

and for the second element

$$\begin{aligned}\text{tr}[\mathbf{E}(\varepsilon_i x'_i) \sum_{t=1}^{T-1} m_{it}^{(2)} p_t p'_t] &= 2\phi \sigma_\varepsilon^2 \sum_{s=2}^T \text{tr}(L'_T H_s A_s H'_s) = 2\phi \sigma_\varepsilon^2 \sum_{s=2}^T \text{tr}(A_s L_s) \\ &= -2\phi \sigma_\varepsilon^2 \left(T - \sum_{s=1}^T \frac{1}{s} \right) = -2\phi T \sigma_\varepsilon^2 + o(1) = O(T).\end{aligned}$$

Hence, we obtain $\mathbf{E}(W' P' M_{Z_l^{(2)}} P \varepsilon) = O(T)$ and the leading bias term (4.7).

For $Z_l = Z_l^{(1)}$, where $m_{it}^{(1)} = \text{tr}(M_{Z_l^{(1)}}) = 2$, the above results in

$$\begin{aligned}\text{tr}[\mathbf{E}(\varepsilon_i y'_{i(-1)}) \sum_{t=1}^{T-1} m_{it}^{(1)} p_t p'_t] &= 2\sigma_\varepsilon^2 \text{tr}[(I_T + \beta \phi L'_T) \Gamma'_T L'_T A_T] = 2\sigma_\varepsilon^2 [\text{tr}(\Pi_T) + \beta \phi \text{tr}(\Pi_T L_T)] \\ \text{tr}[\mathbf{E}(\varepsilon_i x'_i) \sum_{t=1}^{T-1} m_{it}^{(1)} p_t p'_t] &= 2\phi \sigma_\varepsilon^2 \text{tr}(L'_T A_T),\end{aligned}$$

giving $\mathbf{E}(W' P' M_{Z_l^{(1)}} P \varepsilon) = O(1)$, and for the leading bias term we then find (4.8).

Next we investigate $\text{GMM}\mathfrak{H}^{(0)}$, and we directly examine

$$\begin{aligned}\mathbf{E}(F' v) &= \mathbf{E}[W^{*'} Z_l^{(0)} (Z_l^{(0)'} Z_l^{(0)})^{-1} Z_l^{(0)'} \varepsilon_i^*] \\ &= \mathbf{E} \left[\left(\sum_{i=1}^N W_i^{*'} Z_{li}^{(0)} \right) \left(\sum_{i=1}^N Z_{li}^{(0)'} Z_{li}^{(0)} \right)^{-1} \left(\sum_{i=1}^N Z_{li}^{(0)'} \varepsilon_i^* \right) \right].\end{aligned}\tag{A.5}$$

We shall not try to derive the leading term of this expectation, but only its order. Note that according to arguments already used above we have $\mathbf{E}(Z_l^{(0)'} Z_l^{(0)}) = O(n)$ and $(Z_l^{(0)'} Z_l^{(0)})^{-1} = O_p(n^{-1})$, and we can approximate it by $(Z_l^{(0)'} Z_l^{(0)})^{-1} = [\mathbf{E}(Z_l^{(0)'} Z_l^{(0)})]^{-1} + o_p(n^{-1})$. Therefore, the order of the expectation (A.5) will be equivalent to the order of

$$n^{-1} \mathbf{E} \left[\left(\sum_{i=1}^N W_i^{*'} Z_{li}^{(0)} \right) \left(\sum_{i=1}^N Z_{li}^{(0)'} \varepsilon_i^* \right) \right] = n^{-1} \sum_{i=1}^N \mathbf{E}[(W_i^{*'} Z_{li}^{(0)}) (Z_{li}^{(0)'} \varepsilon_i^*)],$$

where the latter equality follows from $\mathbf{E}(Z_{li}^{(0)'} \varepsilon_i^*) = 0$ and the independence of the N individuals. The two elements of the vector $\mathbf{E}[(W_i^{*'} Z_{li}^{(0)})(Z_{li}^{(0)'} \varepsilon_i^*)]$ consist each of the sum of $m^{(0)} = 4$ contributions. Each of these contributions is at most $O(T)$. We shall illustrate that for one such term. We consider the one for the regressor $y_{i(-1)}^*$ and the instrument $x_{i(-1)}$, which both have $T - 1$ elements. We find (upon omitting the subscripts i , and the subscript T from the matrices L , Γ and P , and T or $T - 1$ from ι)

$$\begin{aligned}
& \mathbf{E}(x'_{(-1)} y^*_{(-1)} x'_{(-1)} \varepsilon^*) \\
&= \mathbf{E}[(\bar{x}'_{(-1)} \bar{y}^*_{(-1)} + \tilde{x}'_{(-1)} \tilde{y}^*_{(-1)} + \bar{x}'_{(-1)} \tilde{y}^*_{(-1)} + \tilde{x}'_{(-1)} \bar{y}^*_{(-1)}) (\bar{x}'_{(-1)} \varepsilon^* + \tilde{x}'_{(-1)} \varepsilon^*)] \\
&= \bar{y}^*_{(-1)} \mathbf{E}(\tilde{x}'_{(-1)} \varepsilon^*) \bar{x}_{(-1)} + \bar{x}'_{(-1)} \mathbf{E}(\tilde{y}^*_{(-1)} \varepsilon^*) \bar{x}_{(-1)} + \mathbf{E}(\tilde{x}'_{(-1)} \tilde{y}^*_{(-1)} \tilde{x}'_{(-1)} \varepsilon^*) \\
&= \phi \bar{y}^*_{(-1)} \mathbf{E}(\varepsilon_{(-2)} \varepsilon' P') \bar{x}_{(-1)} + \bar{x}'_{(-1)} \mathbf{E}[PL\Gamma(I + \beta\phi L) \varepsilon \varepsilon' P'] \bar{x}_{(-1)} \\
&\quad + \mathbf{E}\{(\phi \varepsilon'_{(-2)} + \pi \eta_i \iota') PL\Gamma(I + \beta\phi L) \varepsilon (\phi \varepsilon'_{(-2)} + \pi \eta_i \iota') P \varepsilon\} \\
&= \sigma_\varepsilon^2 \phi \bar{y}^*_{(-1)} J L^2 P' \bar{x}_{(-1)} + \sigma_\varepsilon^2 \bar{x}'_{(-1)} PL\Gamma(I + \beta\phi L) P' \bar{x}_{(-1)} \\
&\quad + \phi^2 \mathbf{E}[\varepsilon'_{(-2)} PL\Gamma(I + \beta\phi L) \varepsilon \varepsilon' P' \varepsilon_{(-2)}] + \pi^2 \mathbf{E}[\eta_i^2 \iota' PL\Gamma(I + \beta\phi L) \varepsilon \varepsilon' P' \iota] \\
&= \sigma_\varepsilon^2 \phi \bar{y}^*_{(-1)} J L^2 P' \bar{x}_{(-1)} + \sigma_\varepsilon^2 \bar{x}'_{(-1)} PL\Gamma(I + \beta\phi L) P' \bar{x}_{(-1)} \\
&\quad + \phi^2 \mathbf{E}[\varepsilon'_{(-2)} PL\Gamma(I + \beta\phi L) \varepsilon \varepsilon' P' \varepsilon_{(-2)}] + \sigma_\varepsilon^2 \sigma_\eta^2 \pi^2 \{\iota' PL\Gamma(I + \beta\phi L) P' \iota\}.
\end{aligned}$$

It is obvious that the first two terms are $O(T)$. Note that P is such that $\iota' PL\Gamma P' \iota$ and $\iota' PL\Gamma LP' \iota$ are both $O(T)$. The remaining expectation is $\mathbf{E}[\varepsilon' L^2 J' PL\Gamma(I + \beta\phi L) \varepsilon \varepsilon' P' J L^2 \varepsilon]$. To evaluate this consider first $\mathbf{E}(\varepsilon' A \varepsilon \varepsilon' B \varepsilon)$, where $A = (a_{ij})$ and $B = (b_{ij})$ are general $T \times T$ matrices, but B is lower-diagonal with zeros on its main diagonal. Then, making use of $\varepsilon_i \sim i.i.d.(0, \sigma_\varepsilon^2)$, we find

$$\begin{aligned}
\mathbf{E}(\varepsilon' A \varepsilon \varepsilon' B \varepsilon) &= \mathbf{E} \left[\left(\sum_{i=1}^T \sum_{j=1}^T a_{ij} \varepsilon_i \varepsilon_j \right) \left(\sum_{k=1}^T \sum_{l=1}^T b_{kl} \varepsilon_k \varepsilon_l \right) \right] \\
&= \mathbf{E} \left(\sum_{i=1}^T \sum_{j=i+1}^T a_{ij} b_{ji} \varepsilon_i^2 \varepsilon_j^2 \right) = \sigma_\varepsilon^4 \sum_{i=1}^T \sum_{j=1}^T a_{ij} b_{ji} = \sigma_\varepsilon^4 \text{tr}(AB),
\end{aligned}$$

hence $\mathbf{E}[\varepsilon' L^2 J' PL\Gamma(I + \beta\phi L) \varepsilon \varepsilon' P' J L^2 \varepsilon] = \sigma_\varepsilon^4 \text{tr}[L^2 J' PL\Gamma(I + \beta\phi L) P' J L^2] = O(T)$. Similar results hold for the other terms, and hence for $\text{GMMfl}^{(0)}$ we find that $\mathbf{E}(F'v) = O(n^{-1}NT) = O(1)$ and thus here the order of the leading term of the bias is $O(n^{-1})$.

Next, we consider the GMMld estimators. Employing the permutation matrix G again for $h = 1, 2$ we have here

$$\mathbf{E}(F'v) = \mathbf{E}(W' J' M_{Z_d} J u) = \sum_{t=2}^T \mathbf{E}[W'_t Z_{dt}^{(h)} (Z_{dt}^{(h)'} Z_{dt}^{(h)})^{-1} Z_{dt}^{(h)'} u_t],$$

where $Z_{dt}^{(h)}$ is a $N \times m_{dt}^{(h)}$ matrix. Writing $M_{Z_d} = Z_{dt} (Z_{dt}' Z_{dt})^{-1} Z_{dt}'$ where $\text{tr}(M_{Z_d}) = m_{dt}^{(h)}$ with $m_{dt}^{(2)} = 2t$ and $m_{dt}^{(1)} = 2$ we find

$$\mathbf{E}(W'_t M_{Z_d} u_t) = \mathbf{E} \begin{pmatrix} y'_{t-1} M_{Z_d} u_t \\ x'_t M_{Z_d} u_t \end{pmatrix} = \mathbf{E} \begin{pmatrix} y'_{t-1} M_{Z_d} \eta \\ x'_t M_{Z_d} \eta \end{pmatrix} + \mathbf{E} \begin{pmatrix} y'_{t-1} M_{Z_d} \varepsilon_t \\ x'_t M_{Z_d} \varepsilon_t \end{pmatrix},$$

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

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

where the last term is again zero, now because \bar{y}_{t-1} , \bar{x}_t and Z_{dt} are independent of η . 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 η and therefore cannot contribute to the expectation,

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

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

$$\mathbf{E}(W' J' M_{Z_d^{(h)}} J u) = \sigma_\eta^2 \begin{pmatrix} \alpha \\ \pi \end{pmatrix} \sum_{t=2}^T m_{dt}^{(h)}. \quad (\text{A.6})$$

Note that $\sum_{t=2}^T m_{dt}^{(2)} = 2 \sum_{t=2}^T t = O(T^2)$ and $\sum_{t=2}^T m_{dt}^{(1)} = 2 \sum_{t=2}^T 1 = O(T)$, leading to the results (4.10) and (4.11). Regarding GMMld⁽⁰⁾ we find

$$\begin{aligned} \mathbf{E}(F' v) &= \mathbf{E}[W' Z_d^{(0)} (Z_d^{(0)'} Z_d^{(0)})^{-1} Z_d^{(0)'} u] \\ &= \mathbf{E} \left[\left(\sum_{i=1}^N W_i' Z_{di}^{(0)} \right) \left(\sum_{i=1}^N Z_{di}^{(0)'} Z_{di}^{(0)} \right)^{-1} \left(\sum_{i=1}^N Z_{di}^{(0)'} (\eta_i \iota_{T-1} + \varepsilon_i) \right) \right], \end{aligned}$$

and its order will be equivalent to the order of

$$n^{-1} \mathbf{E} \left[\left(\sum_{i=1}^N W_i' Z_{di}^{(0)} \right) \left(\sum_{i=1}^N Z_{di}^{(0)'} (\eta_i \iota_{T-1} + \varepsilon_i) \right) \right] = n^{-1} \sum_{i=1}^N \mathbf{E} \{ W_i' Z_{di}^{(0)} [Z_{di}^{(0)'} (\eta_i \iota_{T-1} + \varepsilon_i)] \}.$$

Using similar arguments as above, now also using that the third moment of the disturbances is finite, it can be shown that this is also $O(1)$, giving (4.12).

Finally we consider GMMs and rewrite (3.14) using (3.13) as

$$\hat{\delta}_{\text{GMMs}} - \delta = (W^* M_{Z_t^{(2)}} W^* + W' J' M_{Z_d^{(1)}} J W)^{-1} (W^* M_{Z_t^{(2)}} \varepsilon^* + W' J' M_{Z_d^{(1)}} J u),$$

where $\mathbf{E}(W_s' M_{Z_s} W_s) = \bar{Q}_{\text{GMMs}} = O(NT)$. Results derived above directly yield

$$\mathbf{E}(F' v) = \mathbf{E}(W^* M_{Z_t^{(2)}} \varepsilon^*) + \mathbf{E}(W' J' M_{Z_d^{(1)}} J u) = O(T),$$

from which (4.13) follows.

B. Details on the simulation design

For a stationary AR(2) processes $z_t = \frac{1}{1 - \alpha_1 \mathcal{L} - \alpha_2 \mathcal{L}^2} \varepsilon_t$, with $\varepsilon_t \sim i.i.d.(0, \sigma_\varepsilon^2)$, we have

$$\mathbf{V}(z_t) = \left(1 - \alpha_2^2 - \frac{1 + \alpha_2}{1 - \alpha_2} \alpha_1^2 \right)^{-1} \sigma_\varepsilon^2; \quad \mathbf{C}(z_t z_{t-1}) = \frac{\alpha_1}{1 - \alpha_2} \mathbf{V}(z_t). \quad (\text{B.1})$$

Hence, for the ARMA(2,1) process $w_t = (1 + \theta \mathcal{L}) z_t$ we have

$$\mathbf{V}(w_t) = (1 + \theta^2) \mathbf{V}(z_t) + 2\theta \mathbf{C}(z_t, z_{t-1}) = \left(1 + \theta^2 + 2\theta \frac{\alpha_1}{1 - \alpha_2} \right) \mathbf{V}(z_t). \quad (\text{B.2})$$

Now note that in model (2.1)

$$y_{it} = \frac{\beta}{1 - \gamma\mathcal{L}}x_{it} + \frac{1}{1 - \gamma}\eta_i + \frac{1}{1 - \gamma\mathcal{L}}\varepsilon_{it}. \quad (\text{B.3})$$

Substitution of (5.2) gives for scheme 1

$$y_{it}^{(1)} = \frac{\beta}{(1 - \gamma\mathcal{L})(1 - \rho_1\mathcal{L})}\xi_{it} + \frac{1 + \beta\pi_1}{1 - \gamma}\eta_i + \frac{1 + \beta\phi_1\mathcal{L}}{1 - \gamma\mathcal{L}}\varepsilon_{it}. \quad (\text{B.4})$$

Hence, using (B.1) and (B.2), we find

$$\mathbf{V}(y_{it}^{(1)}) = \sigma_\xi^2 \frac{\beta^2(1 + \gamma\rho_1)}{(1 - \rho_1^2)(1 - \gamma^2)(1 - \gamma\rho_1)} + \sigma_\varepsilon^2 \frac{1 + 2\gamma\beta\phi_1 + \beta^2\phi_1^2}{1 - \gamma^2} + \sigma_\eta^2 \left(\frac{1 + \beta\pi_1}{1 - \gamma} \right)^2. \quad (\text{B.5})$$

This yields for the signal-to-noise ratio

$$\zeta^{(1)} = \sigma_\xi^2 \beta^2 \frac{1 + \gamma\rho_1}{(1 - \rho_1^2)(1 - \gamma^2)(1 - \gamma\rho_1)} + \frac{(\gamma + \beta\phi_1)^2}{1 - \gamma^2} \quad (\text{B.6})$$

from which (5.11) easily follows. From (B.5) it is also found that, in order to achieve contributions to $\mathbf{V}(y_{it}^{(1)})$ from the variance components η_i and ε_{it} in a proportion with ratio μ^2 , one has to choose

$$\sigma_\eta^2 \left(\frac{1 + \beta\pi_1}{1 - \gamma} \right)^2 = \mu^2 \sigma_\varepsilon^2 \left(\frac{1 + 2\gamma\beta\phi_1 + \beta^2\phi_1^2}{1 - \gamma^2} \right), \quad (\text{B.7})$$

from which (5.8) follows. According to (5.4) scheme 2 implies $(1 - \rho_2\mathcal{L})x_{it}^{(2)} = \phi_2\mathcal{L}y_{it}^{(2)} + \pi_2\eta_i + \xi_{it}$, and substitution in model (2.1) yields $(1 - \alpha_1\mathcal{L} - \alpha_2\mathcal{L}^2)y_{it}^{(2)} = \beta\xi_{it} + (1 - \rho_2 + \beta\pi_2)\eta_i + (1 - \rho_2\mathcal{L})\varepsilon_{it}$, with $\alpha_1 = \gamma + \beta\phi_2 + \rho_2$ and $\alpha_2 = -\gamma\rho_2$. Hence, we now have

$$y_{it}^{(2)} = \frac{1}{1 - \alpha_1\mathcal{L} - \alpha_2\mathcal{L}^2} [\beta\xi_{it} + (1 - \rho_2 + \beta\pi_2)\eta_i + (1 - \rho_2\mathcal{L})\varepsilon_{it}], \quad (\text{B.8})$$

from which we obtain the results

$$\mathbf{V}[y_{it}^{(2)}] = \left[\sigma_\xi^2 \beta^2 + \sigma_\varepsilon^2 \left(1 + \rho_2^2 - 2\rho_2 \frac{\gamma + \beta\phi_2 + \rho_2}{1 + \gamma\rho_2} \right) \right] \times \quad (\text{B.9})$$

$$\left[1 - \gamma^2 \rho_2^2 - \frac{1 - \gamma\rho_2}{1 + \gamma\rho_2} (\gamma + \beta\phi_2 + \rho_2)^2 \right]^{-1} + \sigma_\eta^2 \left[\frac{1 - \rho_2 + \beta\pi_2}{(1 - \gamma)(1 - \rho_2) - \beta\phi_2} \right]^2,$$

$$\zeta^{(2)} = \left[\sigma_\xi^2 \beta^2 + \sigma_\varepsilon^2 \left(1 + \rho_2^2 - 2\rho_2 \frac{\gamma + \beta\phi_2 + \rho_2}{1 + \gamma\rho_2} \right) \right] \times$$

$$\left[1 - \gamma^2 \rho_2^2 - \frac{1 - \gamma\rho_2}{1 + \gamma\rho_2} (\gamma + \beta\phi_2 + \rho_2)^2 \right]^{-1} - 1, \quad (\text{B.10})$$

and (5.12) 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

$$\sigma_\eta^2 \left[\frac{1 - \rho_2 + \beta\pi_2}{(1 - \gamma)(1 - \rho_2) - \beta\phi_2} \right]^2 = \mu^2 \sigma_\varepsilon^2 \left(1 + \rho_2^2 - 2\rho_2 \frac{\gamma + \beta\phi_2 + \rho_2}{1 + \gamma\rho_2} \right) \times \quad (\text{B.11})$$

$$\left[1 - \gamma^2 \rho_2^2 - \frac{1 - \gamma\rho_2}{1 + \gamma\rho_2} (\gamma + \beta\phi_2 + \rho_2)^2 \right]^{-1},$$

which yields (5.9).

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

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

	further restrictions on $N \geq 2$ and $T \geq 2$	equation formulated in:			bias affected by $\sigma_\eta^2/\sigma_\varepsilon^2$ and by π
		levels	forward orth. dev.	first differences	
<i>LSDV</i>	none	$O(N^0T^{-1})$	$O(N^0T^{-1})$	$O(N^0T^{-1})$	no
<i>GLS</i>	none	$O(N^0T^{-1})$	–	–	heavily
<i>GMMfl</i> ⁽²⁾	$N \geq K(T - 1)$	–	$O(N^{-1}T^0)$	$O(N^{-1}T^0)$	yes
<i>GMMfl</i> ⁽¹⁾	none	–	$O(N^{-1}T^{-1})$	–	yes
<i>GMMfl</i> ⁽⁰⁾	none	–	$O(N^{-1}T^{-1})$	–	yes
<i>GMMld</i> ⁽²⁾	$N \geq K(T - 1)$	$O(N^{-1}T)$	–	–	heavily
<i>GMMld</i> ⁽¹⁾	none	$O(N^{-1}T^0)$	–	–	heavily
<i>GMMld</i> ⁽⁰⁾	none	$O(N^{-1}T^{-1})$	–	–	heavily
<i>GMMs</i>	$N \geq K(T - 1)$	$\underbrace{\hspace{10em}}_{\text{combining all in a hybrid system}}$ $O(N^{-1}T^0)$			heavily

” – ” means: is not defined or not relevant; K is number of regressors in model

Table 2: Simulation results for $T = 10$, $N = 20$, scheme 1, disturbances normal

<i>design</i>	<i>estimator</i>	\bar{B}_γ	\bar{B}_β	<i>bias</i> γ	<i>bias</i> β	<i>ster</i> γ	<i>ster</i> β	<i>rmse</i> γ	<i>rmse</i> β	
1	<i>LSDV</i>	-0.13	0.00	-0.14	0.00	0.06	0.04	0.15	0.04	
	<i>GLS</i>	0.02	-0.00	0.02	-0.00	0.04	0.03	0.04	0.03	
	$\gamma = 0.75$	<i>FGLS</i>			0.12	-0.03	0.04	0.04	0.13	0.05
	$\zeta = 3$	<i>GMMfl</i> ⁽²⁾	-0.11	-0.01	-0.14	-0.01	0.08	0.04	0.16	0.04
	$\pi = 0$	<i>GMMfl</i> ⁽¹⁾	-0.04	-0.01	-0.05	-0.01	0.10	0.05	0.11	0.05
	$\phi = 0$	<i>GMMfl</i> ⁽⁰⁾			-0.01	-0.00	0.11	0.05	0.11	0.05
	$\mu = 1$	<i>GMMld</i> ⁽²⁾	0.09	-0.01	0.07	-0.01	0.05	0.04	0.08	0.04
	$\rho = 0.5$	<i>GMMld</i> ⁽¹⁾	0.03	0.00	0.05	-0.00	0.08	0.05	0.09	0.05
		<i>GMMld</i> ⁽⁰⁾			-0.01	-0.00	0.10	0.05	0.10	0.05
		<i>GMMs</i>	-0.05	-0.00	-0.05	-0.00	0.06	0.04	0.08	0.04
2	<i>LSDV</i>	-0.19	0.06	-0.20	0.06	0.07	0.15	0.21	0.16	
	<i>GLS</i>	0.03	-0.02	0.02	-0.02	0.04	0.08	0.05	0.08	
	$\gamma = 0.75$	<i>FGLS</i>			0.15	-0.13	0.04	0.06	0.16	0.15
	$\zeta = 3$	<i>GMMfl</i> ⁽²⁾	-0.17	0.03	-0.22	0.03	0.09	0.23	0.24	0.23
	$\pi = 0$	<i>GMMfl</i> ⁽¹⁾	-0.06	0.00	-0.10	-0.01	0.14	0.40	0.18	0.40
	$\phi = 0$	<i>GMMfl</i> ⁽⁰⁾			-0.02	-0.00	0.16	0.58	0.16	0.58
	$\mu = 1$	<i>GMMld</i> ⁽²⁾	0.12	-0.09	0.09	-0.07	0.05	0.09	0.10	0.11
	$\rho = 0.95$	<i>GMMld</i> ⁽¹⁾	0.03	-0.02	0.06	-0.04	0.09	0.18	0.11	0.18
		<i>GMMld</i> ⁽⁰⁾			-0.02	-0.01	0.14	0.43	0.14	0.43
		<i>GMMs</i>	-0.07	0.03	-0.08	0.03	0.07	0.15	0.10	0.15
3	<i>LSDV</i>	-0.19	0.06	-0.20	0.06	0.07	0.15	0.21	0.16	
	<i>GLS</i>	0.00	0.00	-0.01	0.01	0.05	0.07	0.05	0.07	
	$\gamma = 0.75$	<i>FGLS</i>			0.05	-0.05	0.05	0.06	0.07	0.08
	$\zeta = 3$	<i>GMMfl</i> ⁽²⁾	-0.14	0.02	-0.18	0.02	0.09	0.22	0.20	0.23
	$\pi = 0$	<i>GMMfl</i> ⁽¹⁾	-0.04	0.00	-0.06	-0.01	0.10	0.38	0.12	0.38
	$\phi = 0$	<i>GMMfl</i> ⁽⁰⁾			-0.02	-0.00	0.12	0.60	0.12	0.60
	$\mu = 0$	<i>GMMld</i> ⁽²⁾	0.00	0.00	-0.01	0.01	0.06	0.09	0.06	0.09
	$\rho = 0.95$	<i>GMMld</i> ⁽¹⁾	0.00	0.00	-0.01	0.00	0.10	0.17	0.10	0.17
		<i>GMMld</i> ⁽⁰⁾			-0.02	-0.01	0.12	0.38	0.12	0.38
		<i>GMMs</i>	-0.09	0.04	-0.11	0.05	0.07	0.14	0.13	0.15
4	<i>LSDV</i>	-0.19	0.06	-0.20	0.06	0.07	0.15	0.21	0.16	
	<i>GLS</i>	0.05	-0.02	0.04	-0.02	0.04	0.12	0.06	0.12	
	$\gamma = 0.75$	<i>FGLS</i>			0.24	-0.21	0.01	0.04	0.24	0.21
	$\zeta = 3$	<i>GMMfl</i> ⁽²⁾	-0.20	0.03	-0.26	0.04	0.10	0.23	0.28	0.23
	$\pi = 0$	<i>GMMfl</i> ⁽¹⁾	-0.16	0.00	-0.27	-0.01	0.23	0.44	0.35	0.44
	$\phi = 0$	<i>GMMfl</i> ⁽⁰⁾			-0.02	0.01	0.38	0.64	0.38	0.64
	$\mu = 5$	<i>GMMld</i> ⁽²⁾	0.28	-0.22	0.23	-0.18	0.02	0.08	0.23	0.19
	$\rho = 0.95$	<i>GMMld</i> ⁽¹⁾	0.11	-0.07	0.22	-0.14	0.04	0.18	0.23	0.23
		<i>GMMld</i> ⁽⁰⁾			0.05	0.02	0.29	0.91	0.30	0.91
		<i>GMMs</i>	0.08	-0.04	0.17	-0.08	0.04	0.15	0.17	0.17
5	<i>LSDV</i>	-0.15	0.04	-0.16	0.05	0.06	0.07	0.17	0.08	
	<i>GLS</i>	0.02	-0.02	0.02	-0.01	0.04	0.04	0.04	0.05	
	$\gamma = 0.75$	<i>FGLS</i>			0.11	-0.09	0.04	0.04	0.11	0.10
	$\zeta = 9$	<i>GMMfl</i> ⁽²⁾	-0.12	0.02	-0.16	0.02	0.08	0.10	0.18	0.11
	$\pi = 0$	<i>GMMfl</i> ⁽¹⁾	-0.04	0.00	-0.06	-0.01	0.11	0.18	0.13	0.18
	$\phi = 0$	<i>GMMfl</i> ⁽⁰⁾			-0.02	-0.00	0.15	0.34	0.15	0.34
	$\mu = 1$	<i>GMMld</i> ⁽²⁾	0.09	-0.07	0.06	-0.05	0.05	0.05	0.08	0.07
	$\rho = 0.95$	<i>GMMld</i> ⁽¹⁾	0.02	-0.01	0.04	-0.03	0.08	0.09	0.09	0.09
		<i>GMMld</i> ⁽⁰⁾			-0.02	-0.00	0.11	0.18	0.11	0.18
		<i>GMMs</i>	-0.05	0.02	-0.05	0.02	0.06	0.07	0.08	0.08
6	<i>LSDV</i>	-0.15	0.04	-0.16	0.05	0.06	0.07	0.17	0.08	
	<i>GLS</i>	0.04	-0.02	0.04	-0.01	0.04	0.06	0.06	0.06	
	$\gamma = 0.75$	<i>FGLS</i>			0.23	-0.19	0.01	0.03	0.23	0.20
	$\zeta = 9$	<i>GMMfl</i> ⁽²⁾	-0.15	0.02	-0.20	0.03	0.09	0.11	0.22	0.11
	$\pi = 0$	<i>GMMfl</i> ⁽¹⁾	-0.11	0.00	-0.20	-0.01	0.20	0.20	0.29	0.20
	$\phi = 0$	<i>GMMfl</i> ⁽⁰⁾			-0.01	0.01	0.25	0.30	0.25	0.30
	$\mu = 5$	<i>GMMld</i> ⁽²⁾	0.27	-0.21	0.23	-0.17	0.02	0.05	0.23	0.18
	$\rho = 0.95$	<i>GMMld</i> ⁽¹⁾	0.11	-0.06	0.21	-0.12	0.04	0.09	0.21	0.15
		<i>GMMld</i> ⁽⁰⁾			0.02	0.02	0.26	0.43	0.26	0.43
		<i>GMMs</i>	0.07	-0.03	0.15	-0.07	0.05	0.08	0.16	0.10

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

<i>design</i>	<i>estimator</i>	\bar{B}_γ	\bar{B}_β	<i>bias</i> γ	<i>bias</i> β	<i>ster</i> γ	<i>ster</i> β	<i>rmse</i> γ	<i>rmse</i> β	
7	<i> LSDV </i>	-0.20	0.05	-0.21	0.05	0.07	0.08	0.22	0.09	
	<i> GLS </i>	0.04	-0.08	0.04	-0.08	0.05	0.07	0.06	0.11	
	$\gamma = 0.75$	<i> FGLS </i>			0.16	-0.14	0.04	0.06	0.17	0.16
	$\zeta = 3$	<i> GMMfl⁽²⁾ </i>	-0.17	0.01	-0.23	0.02	0.10	0.08	0.25	0.09
	$\pi = 0$	<i> GMMfl⁽¹⁾ </i>	-0.08	0.01	-0.11	0.01	0.15	0.11	0.18	0.11
	$\phi = 1$	<i> GMMfl⁽⁰⁾ </i>			-0.02	0.01	0.15	0.10	0.15	0.10
	$\mu = 1$	<i> GMMld⁽²⁾ </i>	0.15	-0.12	0.12	-0.10	0.05	0.08	0.13	0.12
	$\rho = 0.95$	<i> GMMld⁽¹⁾ </i>	0.05	-0.03	0.08	-0.05	0.09	0.09	0.12	0.10
		<i> GMMld⁽⁰⁾ </i>			-0.01	-0.00	0.12	0.10	0.12	0.10
		<i> GMMs </i>	-0.05	-0.01	-0.05	-0.01	0.08	0.08	0.09	0.09
8	<i> LSDV </i>	-0.19	-0.03	-0.20	-0.03	0.09	0.08	0.22	0.09	
	<i> GLS </i>	0.03	0.01	0.02	0.02	0.04	0.04	0.05	0.05	
	$\gamma = 0.75$	<i> FGLS </i>			0.13	-0.09	0.04	0.05	0.13	0.10
	$\zeta = 3$	<i> GMMfl⁽²⁾ </i>	-0.17	-0.04	-0.23	-0.06	0.14	0.11	0.27	0.13
	$\pi = 0$	<i> GMMfl⁽¹⁾ </i>	-0.09	-0.05	-0.14	-0.08	0.25	0.21	0.29	0.22
	$\phi = -1$	<i> GMMfl⁽⁰⁾ </i>			-0.03	-0.02	0.28	0.20	0.28	0.20
	$\mu = 1$	<i> GMMld⁽²⁾ </i>	0.08	-0.01	0.06	-0.00	0.05	0.05	0.08	0.06
	$\rho = 0.95$	<i> GMMld⁽¹⁾ </i>	0.02	0.01	0.04	0.01	0.10	0.09	0.11	0.09
		<i> GMMld⁽⁰⁾ </i>			-0.01	-0.00	0.22	0.16	0.22	0.16
		<i> GMMs </i>	-0.05	0.01	-0.06	0.01	0.08	0.07	0.10	0.08
9	<i> LSDV </i>	-0.19	0.06	-0.20	0.06	0.07	0.15	0.21	0.16	
	<i> GLS </i>	0.03	0.01	0.02	0.02	0.04	0.07	0.05	0.08	
	$\gamma = 0.75$	<i> FGLS </i>			0.13	-0.09	0.04	0.06	0.14	0.11
	$\zeta = 3$	<i> GMMfl⁽²⁾ </i>	-0.16	0.02	-0.21	0.02	0.09	0.23	0.23	0.23
	$\pi = 1$	<i> GMMfl⁽¹⁾ </i>	-0.05	-0.01	-0.09	-0.03	0.13	0.40	0.16	0.40
	$\phi = 0$	<i> GMMfl⁽⁰⁾ </i>			-0.02	-0.01	0.15	0.64	0.16	0.64
	$\mu = 1$	<i> GMMld⁽²⁾ </i>	0.09	-0.03	0.06	-0.02	0.05	0.09	0.08	0.10
	$\rho = 0.95$	<i> GMMld⁽¹⁾ </i>	0.03	-0.00	0.04	-0.00	0.09	0.18	0.10	0.18
		<i> GMMld⁽⁰⁾ </i>			-0.02	-0.01	0.14	0.43	0.14	0.43
		<i> GMMs </i>	-0.07	0.05	-0.08	0.06	0.07	0.15	0.11	0.16
10	<i> LSDV </i>	-0.19	0.06	-0.20	0.06	0.07	0.15	0.21	0.16	
	<i> GLS </i>	0.02	-0.09	0.01	-0.09	0.05	0.08	0.05	0.12	
	$\gamma = 0.75$	<i> FGLS </i>			0.18	-0.19	0.03	0.05	0.18	0.20
	$\zeta = 3$	<i> GMMfl⁽²⁾ </i>	-0.17	0.04	-0.23	0.05	0.10	0.23	0.25	0.23
	$\pi = -1$	<i> GMMfl⁽¹⁾ </i>	-0.07	0.03	-0.13	0.04	0.16	0.41	0.21	0.42
	$\phi = 0$	<i> GMMfl⁽⁰⁾ </i>			-0.02	0.00	0.17	0.59	0.17	0.59
	$\mu = 1$	<i> GMMld⁽²⁾ </i>	0.15	-0.18	0.11	-0.15	0.05	0.08	0.12	0.17
	$\rho = 0.95$	<i> GMMld⁽¹⁾ </i>	0.05	-0.06	0.08	-0.12	0.09	0.17	0.12	0.21
		<i> GMMld⁽⁰⁾ </i>			-0.02	0.01	0.14	0.55	0.15	0.55
		<i> GMMs </i>	-0.06	-0.01	-0.06	-0.04	0.07	0.15	0.09	0.15
11	<i> LSDV </i>	-0.20	0.05	-0.21	0.05	0.07	0.08	0.22	0.09	
	<i> GLS </i>	0.04	-0.03	0.03	-0.03	0.04	0.07	0.06	0.08	
	$\gamma = 0.75$	<i> FGLS </i>			0.14	-0.08	0.04	0.07	0.14	0.10
	$\zeta = 3$	<i> GMMfl⁽²⁾ </i>	-0.17	0.01	-0.22	0.02	0.10	0.08	0.24	0.09
	$\pi = 1$	<i> GMMfl⁽¹⁾ </i>	-0.07	0.00	-0.09	-0.00	0.13	0.11	0.16	0.11
	$\phi = 1$	<i> GMMfl⁽⁰⁾ </i>			-0.02	0.01	0.14	0.10	0.14	0.10
	$\mu = 1$	<i> GMMld⁽²⁾ </i>	0.11	-0.06	0.09	-0.05	0.05	0.08	0.10	0.09
	$\rho = 0.95$	<i> GMMld⁽¹⁾ </i>	0.04	-0.02	0.06	-0.03	0.09	0.09	0.11	0.10
		<i> GMMld⁽⁰⁾ </i>			-0.01	-0.00	0.12	0.10	0.12	0.10
		<i> GMMs </i>	-0.06	0.00	-0.07	0.00	0.08	0.08	0.10	0.08
12	<i> LSDV </i>	-0.20	0.05	-0.21	0.05	0.07	0.08	0.22	0.09	
	<i> GLS </i>	0.06	-0.06	0.06	-0.06	0.04	0.08	0.07	0.10	
	$\gamma = 0.75$	<i> FGLS </i>			0.22	-0.13	0.01	0.06	0.22	0.15
	$\zeta = 3$	<i> GMMfl⁽²⁾ </i>	-0.20	0.01	-0.26	0.02	0.10	0.08	0.28	0.09
	$\pi = 1$	<i> GMMfl⁽¹⁾ </i>	-0.16	0.01	-0.22	0.01	0.20	0.11	0.29	0.11
	$\phi = 1$	<i> GMMfl⁽⁰⁾ </i>			-0.02	0.01	0.31	0.12	0.31	0.12
	$\mu = 5$	<i> GMMld⁽²⁾ </i>	0.25	-0.13	0.21	-0.11	0.02	0.07	0.21	0.13
	$\rho = 0.95$	<i> GMMld⁽¹⁾ </i>	0.10	-0.04	0.20	-0.09	0.04	0.09	0.20	0.12
		<i> GMMld⁽⁰⁾ </i>			0.11	-0.04	0.18	0.11	0.21	0.12
		<i> GMMs </i>	0.08	-0.05	0.16	-0.08	0.05	0.08	0.16	0.12

Table 4: Simulation results for $T = 10$, $N = 20$, scheme 1, disturbances normal

<i>design</i>	<i>estimator</i>	\bar{B}_γ	\bar{B}_β	<i>bias</i> γ	<i>bias</i> β	<i>ster</i> γ	<i>ster</i> β	<i>rmse</i> γ	<i>rmse</i> β	
13	<i>LSDV</i>	-0.10	0.06	-0.11	0.07	0.07	0.12	0.13	0.14	
	<i>GLS</i>	0.01	-0.01	0.01	-0.01	0.06	0.09	0.06	0.09	
	$\gamma = 0.25$	<i>FGLS</i>			0.47	-0.46	0.09	0.10	0.48	0.47
	$\zeta = 3$	<i>GMMfl</i> ⁽²⁾	-0.09	0.04	-0.11	0.05	0.08	0.17	0.13	0.18
	$\pi = 0$	<i>GMMfl</i> ⁽¹⁾	-0.02	0.01	-0.04	0.01	0.11	0.30	0.11	0.30
	$\phi = 0$	<i>GMMfl</i> ⁽⁰⁾			-0.00	0.00	0.10	0.44	0.10	0.44
	$\mu = 1$	<i>GMMld</i> ⁽²⁾	0.29	-0.26	0.22	-0.19	0.09	0.13	0.23	0.23
	$\rho = 0.95$	<i>GMMld</i> ⁽¹⁾	0.05	-0.04	0.09	-0.07	0.10	0.18	0.14	0.20
		<i>GMMld</i> ⁽⁰⁾			-0.01	0.00	0.11	0.43	0.11	0.43
		<i>GMMs</i>	-0.02	0.02	-0.01	0.01	0.08	0.14	0.08	0.14
14	<i>LSDV</i>	-0.10	0.06	-0.11	0.07	0.07	0.12	0.13	0.14	
	<i>GLS</i>	0.01	-0.01	0.01	-0.01	0.06	0.11	0.06	0.11	
	$\gamma = 0.25$	<i>FGLS</i>			0.73	-0.70	0.01	0.04	0.73	0.70
	$\zeta = 3$	<i>GMMfl</i> ⁽²⁾	-0.10	0.05	-0.12	0.06	0.08	0.18	0.15	0.18
	$\pi = 0$	<i>GMMfl</i> ⁽¹⁾	-0.08	0.03	-0.14	0.05	0.20	0.33	0.24	0.33
	$\phi = 0$	<i>GMMfl</i> ⁽⁰⁾			0.01	0.00	0.17	0.42	0.17	0.42
	$\mu = 5$	<i>GMMld</i> ⁽²⁾	0.83	-0.74	0.68	-0.60	0.04	0.10	0.68	0.61
	$\rho = 0.95$	<i>GMMld</i> ⁽¹⁾	0.30	-0.22	0.58	-0.41	0.09	0.23	0.58	0.47
		<i>GMMld</i> ⁽⁰⁾			0.01	0.03	0.25	1.38	0.25	1.38
		<i>GMMs</i>	0.22	-0.14	0.42	-0.27	0.10	0.20	0.44	0.34
15	<i>LSDV</i>	-0.12	0.02	-0.12	0.02	0.07	0.09	0.14	0.09	
	<i>GLS</i>	0.01	-0.07	0.01	-0.07	0.07	0.08	0.07	0.11	
	$\gamma = 0.25$	<i>FGLS</i>			0.54	-0.52	0.07	0.09	0.55	0.53
	$\zeta = 3$	<i>GMMfl</i> ⁽²⁾	-0.09	-0.01	-0.12	-0.00	0.08	0.09	0.14	0.09
	$\pi = 0$	<i>GMMfl</i> ⁽¹⁾	-0.03	0.00	-0.05	0.00	0.12	0.13	0.13	0.13
	$\phi = 1$	<i>GMMfl</i> ⁽⁰⁾			-0.00	0.00	0.10	0.10	0.10	0.10
	$\mu = 1$	<i>GMMld</i> ⁽²⁾	0.44	-0.39	0.35	-0.31	0.09	0.10	0.36	0.33
	$\rho = 0.95$	<i>GMMld</i> ⁽¹⁾	0.09	-0.06	0.17	-0.10	0.12	0.11	0.21	0.15
		<i>GMMld</i> ⁽⁰⁾			0.00	-0.01	0.11	0.11	0.11	0.11
		<i>GMMs</i>	-0.00	-0.03	0.03	-0.06	0.09	0.09	0.09	0.11
16	<i>LSDV</i>	0.01	0.07	0.01	0.08	0.07	0.07	0.07	0.11	
	<i>GLS</i>	0.05	0.06	0.05	0.06	0.05	0.06	0.07	0.09	
	$\gamma = 0.25$	<i>FGLS</i>			0.33	-0.23	0.09	0.11	0.34	0.25
	$\zeta = 3$	<i>GMMfl</i> ⁽²⁾	0.02	0.08	0.03	0.10	0.09	0.11	0.09	0.15
	$\pi = 0$	<i>GMMfl</i> ⁽¹⁾	0.02	0.03	0.03	0.06	0.15	0.19	0.15	0.20
	$\phi = -1$	<i>GMMfl</i> ⁽⁰⁾			0.01	0.02	0.16	0.23	0.16	0.23
	$\mu = 1$	<i>GMMld</i> ⁽²⁾	0.17	-0.03	0.14	-0.02	0.07	0.09	0.16	0.09
	$\rho = 0.95$	<i>GMMld</i> ⁽¹⁾	0.04	0.02	0.07	0.05	0.10	0.13	0.12	0.14
		<i>GMMld</i> ⁽⁰⁾			0.01	0.01	0.16	0.24	0.16	0.24
		<i>GMMs</i>	0.03	0.05	0.05	0.07	0.07	0.09	0.09	0.11
17	<i>LSDV</i>	-0.10	0.06	-0.11	0.07	0.07	0.12	0.13	0.14	
	<i>GLS</i>	0.01	0.04	0.01	0.04	0.06	0.08	0.06	0.09	
	$\gamma = 0.25$	<i>FGLS</i>			0.30	-0.25	0.10	0.11	0.32	0.28
	$\zeta = 3$	<i>GMMfl</i> ⁽²⁾	-0.08	0.04	-0.10	0.04	0.07	0.17	0.13	0.18
	$\pi = 1$	<i>GMMfl</i> ⁽¹⁾	-0.02	0.00	-0.03	0.00	0.09	0.30	0.09	0.30
	$\phi = 0$	<i>GMMfl</i> ⁽⁰⁾			-0.01	0.00	0.10	0.48	0.10	0.48
	$\mu = 1$	<i>GMMld</i> ⁽²⁾	0.12	-0.05	0.09	-0.03	0.08	0.11	0.12	0.11
	$\rho = 0.95$	<i>GMMld</i> ⁽¹⁾	0.02	0.01	0.04	0.02	0.09	0.16	0.10	0.16
		<i>GMMld</i> ⁽⁰⁾			-0.01	-0.00	0.10	0.36	0.10	0.36
		<i>GMMs</i>	-0.04	0.04	-0.04	0.06	0.07	0.13	0.08	0.14
18	<i>LSDV</i>	-0.10	0.06	-0.11	0.07	0.07	0.12	0.13	0.14	
	<i>GLS</i>	-0.03	-0.16	-0.04	-0.16	0.06	0.10	0.07	0.19	
	$\gamma = 0.25$	<i>FGLS</i>			0.65	-0.75	0.05	0.02	0.65	0.75
	$\zeta = 3$	<i>GMMfl</i> ⁽²⁾	-0.09	0.06	-0.12	0.07	0.08	0.18	0.14	0.19
	$\pi = -1$	<i>GMMfl</i> ⁽¹⁾	-0.05	0.07	-0.09	0.11	0.17	0.38	0.19	0.40
	$\phi = 0$	<i>GMMfl</i> ⁽⁰⁾			0.01	-0.03	0.16	0.59	0.16	0.59
	$\mu = 1$	<i>GMMld</i> ⁽²⁾	0.60	-0.88	0.47	-0.73	0.08	0.06	0.48	0.74
	$\rho = 0.95$	<i>GMMld</i> ⁽¹⁾	0.15	-0.35	0.27	-0.70	0.12	0.12	0.30	0.71
		<i>GMMld</i> ⁽⁰⁾			0.01	-0.25	0.22	1.11	0.22	1.14
		<i>GMMs</i>	0.06	-0.31	0.13	-0.60	0.09	0.13	0.16	0.62

Table 5: Simulation results for design 11, $N = 50$, scheme 1, disturbances normal

T	<i>estimator</i>	\bar{B}_γ	\bar{B}_β	<i>bias</i> γ	<i>bias</i> β	<i>ster</i> γ	<i>ster</i> β	<i>rmse</i> γ	<i>rmse</i> β
5	<i> LSDV </i>	-0.41	0.05	-0.41	0.05	0.08	0.08	0.42	0.09
	<i> GLS </i>	0.08	-0.07	0.07	-0.07	0.04	0.06	0.08	0.09
	<i> FGLS </i>			0.11	-0.07	0.03	0.06	0.11	0.09
	<i> GMMfl⁽²⁾ </i>	-0.12	-0.01	-0.20	-0.02	0.16	0.09	0.26	0.09
	<i> GMMfl⁽¹⁾ </i>	-0.09	-0.01	-0.10	-0.01	0.20	0.11	0.22	0.11
	<i> GMMfl⁽⁰⁾ </i>			-0.03	0.00	0.21	0.11	0.21	0.11
	<i> GMMld⁽²⁾ </i>	0.07	-0.03	0.04	-0.02	0.08	0.09	0.09	0.09
	<i> GMMld⁽¹⁾ </i>	0.02	-0.01	0.03	-0.02	0.12	0.09	0.12	0.10
	<i> GMMld⁽⁰⁾ </i>			-0.01	-0.00	0.12	0.09	0.12	0.09
	<i> GMMs </i>	-0.01	-0.01	-0.03	-0.01	0.10	0.09	0.10	0.09
10	<i> LSDV </i>	-0.20	0.05	-0.21	0.05	0.04	0.05	0.21	0.07
	<i> GLS </i>	0.04	-0.03	0.04	-0.03	0.03	0.05	0.05	0.06
	<i> FGLS </i>			0.14	-0.08	0.02	0.04	0.14	0.09
	<i> GMMfl⁽²⁾ </i>	-0.10	0.00	-0.13	0.01	0.07	0.06	0.15	0.06
	<i> GMMfl⁽¹⁾ </i>	-0.03	0.00	-0.04	-0.00	0.09	0.07	0.10	0.07
	<i> GMMfl⁽⁰⁾ </i>			-0.01	0.00	0.09	0.06	0.09	0.06
	<i> GMMld⁽²⁾ </i>	0.08	-0.04	0.06	-0.03	0.04	0.05	0.07	0.06
	<i> GMMld⁽¹⁾ </i>	0.02	-0.01	0.04	-0.02	0.07	0.06	0.08	0.06
	<i> GMMld⁽⁰⁾ </i>			-0.00	0.00	0.07	0.06	0.07	0.06
	<i> GMMs </i>	-0.04	0.00	-0.04	0.00	0.06	0.06	0.07	0.06
20	<i> LSDV </i>	-0.10	0.03	-0.10	0.03	0.03	0.03	0.10	0.05
	<i> GLS </i>	0.01	-0.01	0.01	-0.01	0.02	0.03	0.02	0.03
	<i> FGLS </i>			0.17	-0.12	0.02	0.03	0.17	0.12
	<i> GMMfl⁽²⁾ </i>	-0.08	0.01	-0.09	0.02	0.03	0.04	0.10	0.04
	<i> GMMfl⁽¹⁾ </i>	-0.01	0.00	-0.02	0.00	0.04	0.04	0.05	0.04
	<i> GMMfl⁽⁰⁾ </i>			-0.00	0.00	0.04	0.04	0.04	0.04
	<i> GMMld⁽²⁾ </i>	0.09	-0.05	0.08	-0.04	0.02	0.04	0.08	0.05
	<i> GMMld⁽¹⁾ </i>	0.02	-0.01	0.04	-0.02	0.05	0.04	0.06	0.05
	<i> GMMld⁽⁰⁾ </i>			-0.00	0.00	0.05	0.04	0.05	0.04
	<i> GMMs </i>	-0.05	0.01	-0.05	0.01	0.03	0.04	0.06	0.04
50	<i> LSDV </i>	-0.04	0.02	-0.04	0.02	0.01	0.02	0.04	0.03
	<i> GLS </i>	0.00	-0.00	0.00	-0.00	0.01	0.02	0.01	0.02
	<i> FGLS </i>			0.20	-0.16	0.01	0.02	0.20	0.16
	<i> GMMfl⁽¹⁾ </i>	-0.00	0.00	-0.01	0.00	0.02	0.02	0.02	0.02
	<i> GMMfl⁽⁰⁾ </i>			-0.00	0.00	0.02	0.02	0.02	0.02
	<i> GMMld⁽¹⁾ </i>	0.02	-0.01	0.04	-0.02	0.03	0.03	0.05	0.03
	<i> GMMld⁽⁰⁾ </i>			-0.00	0.00	0.03	0.03	0.03	0.03

Table 6: Simulation results for $T = 10$, $N = 20$, scheme 2, disturbances normal

<i>design</i>	<i>estimator</i>	<i>bias</i> γ	<i>bias</i> β	<i>ster</i> γ	<i>ster</i> β	<i>rmse</i> γ	<i>rmse</i> β
11	<i> LSDV </i>	-0.21	0.03	0.07	0.17	0.22	0.18
	<i> GLS </i>	0.02	0.01	0.04	0.08	0.05	0.08
$\gamma = 0.75$	<i> FGLS </i>	0.14	-0.11	0.04	0.07	0.15	0.13
$\zeta = 3$	<i> GMMfl⁽²⁾ </i>	-0.22	-0.02	0.09	0.26	0.24	0.27
$\pi = 1$	<i> GMMfl⁽¹⁾ </i>	-0.10	-0.07	0.14	0.48	0.17	0.49
$\phi = 1$	<i> GMMfl⁽⁰⁾ </i>	-0.02	-0.02	0.17	0.83	0.17	0.83
$\mu = 1$	<i> GMMld⁽²⁾ </i>	0.06	-0.03	0.06	0.10	0.08	0.10
$\rho = 0.95$	<i> GMMld⁽¹⁾ </i>	0.04	-0.00	0.10	0.20	0.11	0.20
	<i> GMMld⁽⁰⁾ </i>	-0.02	0.01	0.14	0.52	0.14	0.52
	<i> GMMs </i>	-0.09	0.06	0.07	0.16	0.11	0.18
12	<i> LSDV </i>	-0.21	0.03	0.07	0.17	0.22	0.18
	<i> GLS </i>	0.03	0.15	0.04	0.13	0.05	0.20
$\gamma = 0.75$	<i> FGLS </i>	0.23	-0.17	0.01	0.06	0.23	0.18
$\zeta = 3$	<i> GMMfl⁽²⁾ </i>	-0.27	-0.02	0.10	0.27	0.29	0.27
$\pi = 1$	<i> GMMfl⁽¹⁾ </i>	-0.26	-0.15	0.22	0.52	0.34	0.54
$\phi = 1$	<i> GMMfl⁽⁰⁾ </i>	-0.02	-0.01	0.75	0.98	0.75	0.98
$\mu = 5$	<i> GMMld⁽²⁾ </i>	0.21	-0.09	0.02	0.10	0.21	0.13
$\rho = 0.95$	<i> GMMld⁽¹⁾ </i>	0.18	-0.02	0.05	0.20	0.19	0.21
	<i> GMMld⁽⁰⁾ </i>	0.02	0.22	0.23	0.78	0.24	0.81
	<i> GMMs </i>	0.11	0.13	0.05	0.17	0.12	0.21