Orthogonal to Backward Mean Transformation for Dynamic Panel Data Models

Gerdie Everaert*

SHERPPA, Ghent University

February 28, 2011

Abstract

The within-groups estimator is inconsistent in dynamic panels with fixed T as the individual sample mean of the lagged dependent variable used in the within transformation is contemporaneously correlated with the idiosyncratic error term. This paper suggests to transform the lagged dependent variable into orthogonal deviations from its individual backward mean, which is contemporaneously uncorrelated with the idiosyncratic error term. As this transformation eliminates the individual effects as $T \to \infty$ but not for T fixed, this alternative estimator is consistent for $T \to \infty$ but inconsistent for $N \to \infty$ and T fixed. The inconsistency for fixed T is shown to be negligibly small, though. Moreover, a Monte Carlo simulation shows that it has superior small sample properties compared to conventional dynamic panel data estimators.

JEL Classification: C13, C15, C23

Keywords: Dynamic panel, Individual effects, Backward mean, Monte Carlo simulation

1 Introduction

The basic problem in dynamic panel data models is that the lagged dependent variable is by construction correlated with the individual effect in the error term. This renders the least-squares (LS) estimator biased and inconsistent. Consistent estimation requires some transformation to eliminate the individual effects. A within transformation wipes out the individual effects by taking deviations from individual sample means, but the resulting within-groups (WG) estimator is inconsistent when the cross-sectional dimension N tends to infinity with the time dimension T fixed (see e.g. Nickell, 1981). Given this inconsistency, the literature focuses mainly on a first difference transformation to eliminate the individual effect while handling the remaining

*Correspondence to: Tweekerkenstraat 2, B-9000 Gent, Belgium. Email: gerdie.everaert@UGent.be. Website: http://www.sherppa.be.

correlation with the (transformed) error term using instrumental variables (IV) and generalised method of moments (GMM) estimators. These estimators are consistent for large N and fixed T (see Arellano and Bond, 1991; Blundell and Bond, 1998) or for large N and large T (see Anderson and Hsiao, 1982; Alvarez and Arellano, 2003). Especially the first-differenced GMM estimator of Arellano and Bond (1991) and the system GMM estimator of Arellano and Bover (1995) and Blundell and Bond (1998) are increasingly popular. Unfortunately, these GMM estimators (i) have a (much) larger standard error compared to the WG estimator (see e.g. Arellano and Bond, 1991; Kiviet, 1995) and (ii) may suffer from a substantial finite sample bias due to weak instrument problems (see Ziliak, 1997; Bun and Kiviet, 2006; Bun and Windmeijer, 2010). In order to avoid these problems, bias-corrections for the WG estimator have been proposed by, among others, Kiviet (1995), Bun (2003), Bun and Carree (2005) and Everaert and Pozzi (2007). The advantage of these estimators is that they reduce the bias of the WG estimator while maintaining its relatively small dispersion. Although these estimators perform remarkably well in most cases, the remaining bias may still be substantial when T is relatively small. Moreover, they are not always that straightforward to implement.

This paper retains a within-type of transformation but suggests an alternative to taking deviations from the individual sample mean. The problem with the sample mean of the lagged dependent variable is that it includes observations for time $t, \ldots, T$, which are all affected by the idiosyncratic error term at time t. When T is fixed, this results in contemporaneous correlation between the within-transformed lagged dependent variable and the idiosyncratic error term, which in turn implies that the WG estimator is inconsistent for $N \to \infty$ and T fixed. As an alternative, we therefore suggest to transform the lagged dependent variable into orthogonal deviations from its individual backward mean, which is contemporaneously uncorrelated with the idiosyncratic error term. We refer to the estimators based on this orthogonal to backward mean transformation as WGob estimators. We consider a model with and without additional exogenous regressors. First, in a model with no additional exogenous regressors, the WGob estimator is obtained as the LS estimator after transforming the model in orthogonal deviations from the backward mean of the lagged dependent variable. Equivalently, it is obtained by (i) adding the backward mean of the lagged dependent variable as a regressor to the model, which then serves as a proxy for the individual effects, or (ii) instrumenting the lagged dependent variable by the orthogonal deviations from its backward mean, which is similar to the Hausman and Taylor (1981) representation of the WG estimator. Second, the Hausman-Taylor representation of the WGob estimator makes it very easy to add exogenous explanatory variables to the model which (i) serve as their own instruments when they are not correlated with the individual effect or (ii) can be instrumented by the deviations from their sample mean when they are correlated with the individual effect.

The WGob estimator is shown to be consistent for $T \to \infty$ but inconsistent for $N \to \infty$ and T fixed. This is due to the fact that for fixed T the individual effects are not exactly wiped out by orthogonalising on the backward mean of the lagged dependent variable. The inconsistency is shown to be negligibly small, though. Moreover, the WGob estimator is consistent for fixed T in the specific cases where (i) $T = 2$; (ii) the AR(1) coefficient is either zero or increases towards unity and (iii) the ratio of the variance of individual effects over the variance of the idiosyncratic error tends to zero. Monte Carlo experiments further show that the small sample properties of the WGob estimator are superior to those of conventional dynamic panel data estimators, i.e. it considerably outperforms conventional estimators in terms of bias, dispersion and inference in the cases where these estimators fail while not performing much worse in all other cases.

The remainder of this paper is organised as follows. Section 2 presents the model and the assumptions. Section 3 motivates the use of orthogonal deviations from the backward mean of the lagged dependent variable by inspection of the source of the Nickell bias. Section 4 analyses the asymptotic properties of the WGob estimator in a model with only a lagged dependent variable. Section 5 extends the model by adding exogenous explanatory variables. Section 6 presents the results of Monte Carlo experiments comparing the finite sample performance of the WGob estimator to a number of standard dynamic panel data estimators. Section 7 concludes.

2 Model and assumptions

Consider a standard dynamic panel data model with individual effects

$$y_{it} = \theta y_{i,t-1} + \alpha_i + \varepsilon_{it}, \qquad i = 1, \ldots, N; \quad t = 1, \ldots, T, \tag{1}$$

where $|\theta| < 1$, y_{it} is the observation on the dependent variable for unit i at time t and $\alpha_i + \varepsilon_{it}$ is the usual decomposition of the error term into the unobserved individual heterogeneity α_i or individual effect and the idiosyncratic disturbance term ε_{it}. For notational convenience we assume y_{i0} is observed. We make the following assumptions:

Assumption A1. *$\varepsilon_{it} \sim i.i.d.\left(0, \sigma_\varepsilon^2\right)$ across i and t and independent of α_i and y_{i0}.*

Assumption A2. *The initial conditions satisfy*

$$y_{i0} = \frac{\alpha_i}{1-\theta} + \eta_{i0}, \qquad (i = 1, \ldots, N),$$

where η_{i0} is independent of α_i and i.i.d. with the steady state distribution of the homogeneous process so that η_{i0} is the infinite weighted sum $\sum_{s=0}^{\infty} \theta^s \varepsilon_{i,-s}$.

Assumption A3. *$\alpha_i \sim i.i.d.\left(0, \sigma_\alpha^2\right)$ across i.*

For the presentation of the estimators below, it is convenient to write model (1) in the form

$$y_i = \theta y_{i,-1} + \alpha_i \iota_T + \varepsilon_i, \tag{2}$$

where $y_i = (y_{i1}, \ldots, y_{iT})'$, $y_{i,-1} = (y_{i0}, \ldots, y_{i,T-1})'$, ι_T is a $T \times 1$ vector of ones and $\varepsilon_i =$

$(\varepsilon_{i1}, \ldots, \varepsilon_{iT})'$. Upon stacking this information on all N cross-sections, i.e $y = (y'_1, \ldots, y'_N)'$, $y_{-1} = \left(y'_{1,-1}, \ldots, y'_{N,-1}\right)'$, $\alpha = (\alpha_1, \ldots, \alpha_N)'$ and $\varepsilon = (\varepsilon'_1, \ldots, \varepsilon'_N)'$, we have

$$y = \theta y_{-1} + D\alpha + \varepsilon, \tag{3}$$

where $D = I_N \otimes \iota_T$ is a $NT \times N$ dummy variable matrix.

3 Intuitive motivation: the Nickell bias revisited

Let the WG operator Q be given by

$$Q = I_N \otimes Q_T, \qquad \text{with} \quad Q_T = I_T - \iota_T \iota'_T / T \ , \tag{4}$$

which is a symmetric and idempotent matrix that transforms the data into deviations from individual specific sample means:

$$Q_T y_i = \widetilde{y}_i = y_i - \iota_T \bar{y}_i, \quad \text{and} \quad Q_T y_{i,-1} = \widetilde{y}_{i,-1} = y_{i,-1} - \iota_T \bar{y}_{i,-1}, \tag{5}$$

where $\bar{y}_i = T^{-1} \sum_{t=1}^{T} y_{it}$ and $\bar{y}_{i,-1} = T^{-1} \sum_{t=1}^{T} y_{i,t-1}$. Since $Q_T \iota_T = 0$, the individual effects are cancelled out by premultiplying model (3) by Q, obtaining

$$\widetilde{y} = \theta \widetilde{y}_{-1} + \widetilde{\varepsilon}, \tag{6}$$

where $\widetilde{y} = (\widetilde{y}'_1, \ldots, \widetilde{y}'_N)'$, $\widetilde{y}_{-1} = \left(\widetilde{y}'_{1,-1}, \ldots, \widetilde{y}'_{N,-1}\right)'$ and $\widetilde{\varepsilon} = (\widetilde{\varepsilon}'_1, \ldots, \widetilde{\varepsilon}'_N)'$ with $\widetilde{\varepsilon}_i = \varepsilon_i - \iota_T \bar{\varepsilon}_i$ and $\bar{\varepsilon}_i = T^{-1} \sum_{t=1}^{T} \varepsilon_{it}$. The least squares estimate of θ in equation (6) defines the WG estimator

$$\widehat{\theta}^{WG} = \left(\widetilde{y}'_{-1} \widetilde{y}_{-1}\right)^{-1} \widetilde{y}'_{-1} \widetilde{y} = \left(y'_{-1} Q y_{-1}\right)^{-1} y'_{-1} Q y, \tag{7}$$

where use is made of Q being symmetric and idempotent.

The WG estimator can also be written as the least squares estimator for θ after transforming the data into scaled deviations from forward (cf. Arellano and Bover, 1995; Alvarez and Arellano, 2003) or backward means. Define the backward mean operator M_T^b as

$$M_T^b = diag\left[1, \frac{1}{2}, \ldots, \frac{1}{T}\right] \times L_T, \tag{8}$$

where L_T is a $T \times T$ lower triangular matrix of ones, i.e. $L_{T,ij} = 1$ for $i \leq j$ and 0 otherwise, such that Q^b

$$Q^b = I_N \otimes Q_T^b, \quad \text{with} \quad Q_T^b = I_T - M_T^b, \tag{9}$$

is the operator that transforms the data into deviations from backward means

$$Q_T^b y_i = \widetilde{y}_i^b = y_i - \bar{y}_i^b, \quad \text{and} \quad Q_T^b y_{i,-1} = \widetilde{y}_{i,-1}^b = y_{i,-1} - \bar{y}_{i,-1}^b. \tag{10}$$

where $\bar{y}_i^b = \left[\bar{y}_{i1}^b, \ldots, \bar{y}_{iT}^b\right]'$ and $\bar{y}_{i,-1}^b = \left[\bar{y}_{i1,-1}^b, \ldots, \bar{y}_{iT,-1}^b\right]'$ with $\bar{y}_{it}^b = t^{-1}\sum_{s=1}^{t} y_{is}$ and $\bar{y}_{it,-1}^b = t^{-1}\sum_{s=0}^{t-1} y_{is}$.[1] As the rows of Q_T^b add up to zero, i.e. $Q_T^b \iota_T = 0$, the individual effects are also canceled out by premultiplying model (3) by CQ^b, obtaining

$$C\widetilde{y}^b = \theta C\widetilde{y}_{-1}^b + C\widetilde{\varepsilon}^b, \tag{11}$$

where $\widetilde{y}^b = \left(\widetilde{y}_1^{b'}, \ldots, \widetilde{y}_N^{b'}\right)'$, $\widetilde{y}_{-1}^b = \left(\widetilde{y}_{1,-1}^{b'}, \ldots, \widetilde{y}_{N,-1}^{b'}\right)'$ and $\widetilde{\varepsilon}^b = \left(\widetilde{\varepsilon}_1^{b'}, \ldots, \widetilde{\varepsilon}_N^{b'}\right)'$ with $\widetilde{\varepsilon}_i^b = \varepsilon_i - \bar{\varepsilon}_i^b$, $\bar{\varepsilon}_i^b = \left(\bar{\varepsilon}_{i1}^b, \ldots, \bar{\varepsilon}_{iT}^b\right)'$ and $\bar{\varepsilon}_{it}^b = t^{-1}\sum_{s=1}^{t} \varepsilon_{is}$. The scale factor $C = I_N \otimes C_T$, with $C_T = diag\left[1, \sqrt{2}, \sqrt{3/2}, \ldots, \sqrt{T/T-1}\right]$, is introduced to ensure that the transformation preserves the orthogonality of the error terms, i.e. if $\text{Var}(\varepsilon_i) = \sigma^2 I_T$ then $C\widetilde{\varepsilon}_i^b$ also has $\text{Var}\big(C\widetilde{\varepsilon}_i^b\big) = \sigma^2 I_T$. The least squares estimate of θ in equation (11)

$$\widehat{\theta}^{WG} = \left(\widetilde{y}_{-1}^{b'} C' C\widetilde{y}_{-1}^b\right)^{-1} \widetilde{y}_{-1}^{b'} C' C\widetilde{y}^b = \left(y_{-1}' Q^{b'} C' C Q^b y_{-1}\right)^{-1} y_{-1}' Q^{b'} C' C Q^b y, \tag{12}$$

equals the WG estimator in (7) as it can easily be verified that $Q^{b'} C' C Q^b = Q$. Note that opposed to Q, CQ^b is not a symmetric and idempotent matrix.

It is well known that $\widehat{\theta}^{WG}$ is consistent for $T \to \infty$ but inconsistent for $N \to \infty$ and T fixed (cf. Nickell, 1981; Anderson and Hsiao, 1981). Inserting (3) in (7) and using $QD = 0$

$$\widehat{\theta}^{WG} = \theta + \left(\widetilde{y}_{-1}' y_{-1}\right)^{-1} \widetilde{y}_{-1}' \left(D\alpha + \varepsilon\right) = \theta + \left(y_{-1}' Q y_{-1}\right)^{-1} y_{-1}' Q\varepsilon, \tag{13}$$

shows that this inconsistency stems from the fact that for fixed T the term $\frac{1}{N} y_{-1}' Q\varepsilon$ does not converge to zero as $N \to \infty$ since the sample mean $\bar{y}_{i,-1}$ used in the within transformation $\widetilde{y}_{i,t-1} = y_{i,t-1} - \bar{y}_{i,-1}$ is contemporaneously correlated with the idiosyncratic error term ε_{it}. Obtaining a consistent LS estimator for $N \to \infty$ requires that the transformation that eliminates the individual effects produces a variable that is contemporaneously uncorrelated with the transformed disturbance ε_{it}. This suggests using backward means in stead of full sample means. However the representation of the WG estimator in (12) shows that this yields exactly the same estimator. Inserting (11) in (12)

$$\widehat{\theta}^{WG} = \theta + \left(\widetilde{y}_{-1}^{b'} C' C\widetilde{y}_{-1}^b\right)^{-1} \widetilde{y}_{-1}^{b'} C' C\widetilde{\varepsilon}^b = \theta + \left(y_{-1}' Q^{b'} C' C Q^b y_{-1}\right)^{-1} y_{-1}' Q^{b'} C' C Q^b \varepsilon, \tag{14}$$

shows that the inconsistency of the WG estimator can also be seen to stem from the correlation between y_{-1} and ε in (scaled) deviation from their backward means.

[1]Note that as according to this definition $\widetilde{y}_{i1}^b$ is zero it is in principle possible to define Q_T^b as a $(T-1)\times(T-1)$ matrix such that observations for $t = 1$ are dropped. However, for the orthogonal deviations presented below these observations are not exactly equal to zero such that for notational convenience they are also not dropped here.

Remark 1. Interestingly, $\widetilde{y}^b_{it,-1} = y_{i,t-1} - \bar{y}^b_{i,-1}$ is not correlated with the error term ε_{it}. Therefore, in line with the IV representation of the WG estimator (see Hausman and Taylor, 1981), we can also use $\widetilde{y}^b_{-1}$ as an instrument for y_{-1}. This IV estimator for θ in (3) is given by

$$\widehat{\theta}^{IV} = \left(\widetilde{y}^{b\prime}_{-1} y_{-1}\right)^{-1} \widetilde{y}^{b\prime}_{-1} y = \theta + \left(y'_{-1} Q^{b\prime} y_{-1}\right)^{-1} y'_{-1} Q^{b\prime} \varepsilon, \tag{15}$$

where use is made of $Q^b \iota_T = 0$ such that $\widetilde{y}^b_{-1} D\alpha = 0$ by construction. As $\widetilde{y}^b_{it,-1}$ and ε_{it} are uncorrelated, $\widehat{\theta}^{IV}$ is a consistent estimator for θ when $N \to \infty$, $T \to \infty$ or $N, T \to \infty$ jointly. However, it suffers from a weak instruments problem as the R^2 of the first step reduced form instrumental variable regression of $y_{i,t-1}$ on the instrument $\widetilde{y}^b_{i,t-1}$ tends to zero when $\left(\frac{\sigma^2_\alpha}{\sigma^2_\varepsilon}\right) \to \infty$ or θ is close to 1.

4 Orthogonalising regressors to backward means

Instead of taking deviations from backward means, define the orthogonal to backward means operator $Q^b_\perp$

$$Q^b_\perp = I_{NT} - \bar{y}^b_{-1} \left(\bar{y}^{b\prime}_{-1} \bar{y}^b_{-1}\right)^{-1} \bar{y}^{b\prime}_{-1}, \tag{16}$$

where $\bar{y}^b_{-1} = \left(\bar{y}^{b\prime}_{1,-1}, \ldots, \bar{y}^{b\prime}_{N,-1}\right)'$ such that $Q^b_\perp$ has the interpretation of a 'residual maker' matrix, i.e. premultiplying by this matrix transforms the data into residuals of an auxiliary regression on $\bar{y}^b_{-1}$. These residuals are by construction orthogonal to $\bar{y}^b_{-1}$. It is easily verified that $Q^b_\perp$ is a symmetric and idempotent matrix. Premultiplying (3) by $Q^b_\perp$ yields

$$\breve{y}^b = \theta \breve{y}^b_{-1} + \breve{\alpha}^b + \breve{\varepsilon}^b, \tag{17}$$

where $\breve{y}^b$, $\breve{y}^b_{-1}$, $\breve{\alpha}^b$ and $\breve{\varepsilon}^b$ are the residuals from the auxiliary regressions of y, y_{-1}, α and ε on $\bar{y}^b_{-1}$. The LS estimator for θ in (17), we shall refer to this as WGob, is given by

$$\widehat{\theta}^{WG}_\perp = \left(y'_{-1} Q^b_\perp y_{-1}\right)^{-1} y'_{-1} Q^b_\perp y = \left(\breve{y}^{b\prime}_{-1} y_{-1}\right)^{-1} \breve{y}^{b\prime}_{-1} y, \tag{18}$$

where use is made of the symmetry and idempotency of $Q^b_\perp$.

Remark 2. Using the Frisch-Waugh-Lovell theorem, the WGob estimator is numerically identical to the LS estimate for the coefficient on y_{-1} in a regression of y on y_{-1} augmented with $\bar{y}^b_{-1}$. This makes the suggested estimator straightforward to apply in practice. Alternatively, the WGob estimator has an instrumental variables (IV) representation where $\breve{y}^b_{i,t-1}$ is used as an instrument for $y_{i,t-1}$ in the original model in (1), i.e. imposing the following moment condition

$$E\left(\breve{y}^b_{i,t-1}\left(\alpha_i + \varepsilon_{it}\right)\right) = 0. \tag{19}$$

Inserting (3) in (18) yields

$$\widehat{\theta}_{\perp}^{WG} = \theta + \left(\breve{y}_{-1}^{b\prime} y_{-1}\right)^{-1} \breve{y}_{-1}^{b\prime} \left(D\alpha + \varepsilon\right) = \theta + \left(y_{-1}^{\prime} Q_{\perp}^{b} y_{-1}\right)^{-1} y_{-1}^{\prime} Q_{\perp}^{b} \left(D\alpha + \varepsilon\right), \tag{20}$$

In contrast to Q^b, the rows of $Q^b_{\perp}$ do not sum to zero such that $Q^b_{\perp}D$ is not zero by construction. This implies that by premultiplying the data by $Q^b_{\perp}$ the individual effects in α are not cancelled out exactly such that the transformed explanatory variable $\breve{y}^b_{-1}$ in the numerator of (20) is potentially correlated with the composite error term $D\alpha + \varepsilon$. This can also be seen by using (16) to rewrite (20) as

$$\widehat{\theta}_{\perp}^{WG} - \theta = \frac{\left(Q_{\perp}^{b} y_{-1}\right)^{\prime} \left(D\alpha + \varepsilon\right)}{\left(Q_{\perp}^{b} y_{-1}\right)^{\prime} \left(Q_{\perp}^{b} y_{-1}\right)} = \frac{\left(y_{-1} - \widehat{\delta}_T \overline{y}_{-1}^{b}\right)^{\prime} \left(D\alpha + \varepsilon\right)}{\left(y_{-1} - \widehat{\delta}_T \overline{y}_{-1}^{b}\right)^{\prime} y_{-1}}, \tag{21}$$

where

$$\widehat{\delta}_T = \frac{\overline{y}_{-1}^{b\prime} y_{-1}}{\overline{y}_{-1}^{b\prime} \overline{y}_{-1}^{b}} = 1 + \frac{\overline{y}_{-1}^{b\prime} \widetilde{y}_{-1}^{b}}{\overline{y}_{-1}^{b\prime} \overline{y}_{-1}^{b}}, \tag{22}$$

which is not necessarily equal to 1 such that the individual effects are not necessarily cancelled out completely.

Let σ_y^2 denote the variance of $y_{i,t-1}$ and, for a given value of T, $\sigma^2_{\breve{y}^b,T}$ the variance of $\breve{y}^b_{i,t-1}$, $\sigma^2_{\overline{y}^b,T}$ the variance of $\overline{y}^b_{i,t-1}$, $\sigma_{y\overline{y}^b,T}$ the covariance between $y_{i,t-1}$ and $\overline{y}^b_{i,t-1}$ and $\sigma_{\overline{y}^b\widetilde{y}^b,T}$ the covariance between $\overline{y}^b_{i,t-1}$ and $\widetilde{y}^b_{i,t-1}$. By virtue of $Q^b_{\perp}y_{-1}$ being predetermined and using straightforward calculations, we have for $N \to \infty$ and T fixed

$$\underset{N\to\infty}{\text{plim}} \left(\widehat{\theta}_{\perp}^{WG} - \theta\right) = \frac{1-\delta_T}{\sigma^2_{\breve{y}^b,T}} \frac{\sigma_\alpha^2}{1-\theta} = \frac{1-\delta_T}{\sigma_y^2 - \delta_T \sigma_{y\overline{y}^b,T}} \frac{\sigma_\alpha^2}{1-\theta}, \tag{23}$$

where $\delta_T = \underset{N\to\infty}{\text{plim}}\, \widehat{\delta}_T = 1 + \sigma_{\overline{y}^b\widetilde{y}^b,T} \Big/ \sigma^2_{\overline{y}^b,T}$. The results collected in the following Lemma are useful to establish the asymptotic properties of $\widehat{\theta}^{WG}_{\perp}$. All proofs are in the appendix.

Lemma 1. *Under assumptions (A1)-(A3):*

$$\sigma^2_{\overline{y}^b,T} = \frac{\sigma_\alpha^2}{(1-\theta)^2} + \frac{\sigma_\varepsilon^2}{(1-\theta)^2} \frac{1}{T} \sum_{t=1}^{T} \frac{1}{t} \left(1 - \frac{2\theta}{t} \frac{1-\theta^t}{1-\theta^2}\right) = \frac{\sigma_\alpha^2}{(1-\theta)^2} + O\left(\frac{\log(T)}{T}\right), \tag{24}$$

$$\sigma_{y\overline{y}^b,T} = \frac{\sigma_\alpha^2}{(1-\theta)^2} + \frac{\sigma_\varepsilon^2}{1-\theta^2} \frac{1}{T} \sum_{t=1}^{T} \frac{1}{t} \frac{1-\theta^t}{1-\theta} = \frac{\sigma_\alpha^2}{(1-\theta)^2} + O\left(\frac{\log T}{T}\right), \tag{25}$$

$$\sigma_{\overline{y}^b\widetilde{y}^b,T} = -\frac{\theta}{1+\theta} \frac{\sigma_\varepsilon^2}{(1-\theta)^2} \frac{1}{T} \sum_{t=1}^{T} \frac{1}{t} \left(1 + \theta^{t-1} - \frac{2}{t} \frac{1-\theta^t}{1-\theta}\right) = O\left(\frac{\log(T)}{T}\right). \tag{26}$$

Using (26) and (24), δ_T is given by

$$\delta_T = \underset{N\to\infty}{\text{plim}}\ \widehat{\delta}_T = 1 + \frac{\sigma_{\overline{y}\widetilde{y}^b,T}}{\sigma^2_{\widetilde{y}^b,T}} = 1 - \frac{\theta}{1+\theta}\frac{\frac{1}{T}\sum_{t=1}^{T}\frac{1}{t}\left(1+\theta^{t-1}-\frac{2}{t}\frac{1-\theta^t}{1-\theta}\right)}{\frac{\sigma^2_\alpha}{\sigma^2_\varepsilon}+\frac{1}{T}\sum_{t=1}^{T}\frac{1}{t}\left(1-\frac{2\theta}{t}\frac{1-\theta^t}{1-\theta^2}\right)}. \tag{27}$$

For large T, this inconsistency has the expansion

$$\delta_T = 1 + O\left(\frac{\log T}{T}\right). \tag{28}$$

The implication of Lemma 1 is that $\widehat{\delta}_T$ differs from 1 as $N \to \infty$ and T fixed but converges to 1 as $T \to \infty$ regardless of the asymptotic behaviour of N (which includes N fixed). From (21) it is clear that this implies that the WGob estimator is inconsistent for $N \to \infty$ and T fixed but consistent for $T \to \infty$. The exact form of the inconsistency and its large T expansion are given in the following Theorem.

Theorem 1. *As $N \to \infty$ and T fixed, the inconsistency of the WGob estimator for θ in model (1) under assumptions A1-A3 is given by*

$$\underset{N\to\infty}{plim}\left(\widehat{\theta}^{WG}_{\perp}-\theta\right) = \frac{\theta\left(1-\theta\right)A_T}{\left(1-\theta\right)+\theta A_T - B_T + \frac{\sigma^2_\varepsilon}{\sigma^2_\alpha}C_T}, \tag{29}$$

$$\text{with} \qquad A_T = \frac{1}{T}\sum_{t=1}^{T}\frac{1}{t}\left(1+\theta^{t-1}-\frac{2}{t}\frac{1-\theta^t}{1-\theta}\right),$$

$$B_T = \frac{1}{T}\sum_{t=1}^{T}\frac{1}{t}\left(1-\theta^t\right),$$

$$C_T = \left(1-\theta\right)\frac{1}{T}\sum_{t=1}^{T}\frac{1}{t}\left(1-\frac{2}{t}\frac{1-\theta^t}{1-\theta^2}\right) - \frac{B_T^2}{1+\theta}.$$

For large T, this inconsistency has the expansion

$$\underset{N\to\infty}{plim}\left(\widehat{\theta}^{WG}_{\perp}-\theta\right) = \theta\frac{\log T}{T}. \tag{30}$$

By direct calculation, some more specific quantitative results implied by Theorem 1 are

(a) $\underset{N\to\infty}{\text{plim}}\left(\widehat{\theta}^{WG}_{\perp}-\theta\right)$ is positive for $0 < \theta < 1$ and negative for $-1 < \theta < 0$.

(b) $\underset{N\to\infty}{\text{plim}}\left(\widehat{\theta}^{WG}_{\perp}-\theta\right)$ increases in $\frac{\sigma^2_\alpha}{\sigma^2_\varepsilon}$ with an upper bound given by

$$\underset{N,\frac{\sigma^2_\alpha}{\sigma^2_\varepsilon}\to\infty}{\text{plim}}\left(\widehat{\theta}^{WG}_{\perp}-\theta\right) = \frac{\theta\left(1-\theta\right)A_T}{\left(1-\theta\right)+\theta A_T - B_T}, \tag{31}$$

(c) $\underset{N\to\infty}{\text{plim}}\left(\widehat{\theta}^{WG}_{\perp}-\theta\right) = 0$ in the following cases (i) $T = 2$, (ii) $\theta = 0$, (iii) $\theta \to 1$ and (iv)

$\sigma_\varepsilon^2 / \sigma_\alpha^2 \to \infty$.

(d) For small values of T the upper bound of the inconsistency is given by

$$\operatorname*{plim}_{N\to\infty} \left(\widehat{\theta}_\perp^{WG} - \theta\right) = 0 \qquad \text{for} \quad T = 2, \tag{32}$$

$$\operatorname*{plim}_{N, \frac{\sigma_\alpha^2}{\sigma_\varepsilon^2}\to\infty} \left(\widehat{\theta}_\perp^{WG} - \theta\right) = \frac{\theta(1-\theta)}{4(3 - 3/8 + \theta)} \qquad \text{for} \quad T = 3. \tag{33}$$

(e) The inconsistency is $O(\log T / T)$ such that, compared to WG, convergence is at a slower rate as T grows large.

Comparing (29)-(33) with the asymptotic bias expressions for the WG estimator (see e.g. Nickell, 1981) shows that, over the relevant range $0 \leq \theta \leq 1$, the (upper bound of the) inconsistency of $\widehat{\theta}_\perp^{WG}$ for fixed T is much smaller than that of the WG estimator. Moreover, $\widehat{\theta}_\perp^{WG}$ is consistent for $N \to \infty$ and fixed T in the specific cases where (i) $T = 2$, (ii) θ is either zero or tends to unity and (iii) $\sigma_\alpha^2 / \sigma_\varepsilon^2 = 0$. This can also be seen from (27) which shows that $\delta_T = 1$ when $T = 2$, $\theta = 0$ or $\theta \to 1$ such that the individual effects are cancelled out completely while for $\sigma_\alpha^2 = 0$ there is no need to eliminate individual effects. Note that in the cases where T is very small or θ is close to 1, standard estimators like GMM and bias-corrected WG estimators are known to fail. Figure 1 plots the upper bound of the inconsistency of the WGob estimator, calculated from (31), for various values of θ and T. The most important conclusion from this graph is that the upper bound on the inconsistency is negligibly small for all values of θ and T, i.e. it is never larger than 0.04. Note that compared to the WG estimator, $\widehat{\theta}_\perp^{WG}$ converges at a slower rate when $T \to \infty$. This slower rate of convergence is due to the fact that in calculating the backward mean only information up to time t is used, i.e. as T grows $\bar{y}_{i,t-1}^b$ is not updated, while the sample mean used to construct the WG estimator uses information up to time T.

5 Extension: inclusion of exogenous variables

Adding explanatory variables to the model in (1) yields

$$y_{it} = \alpha_i + \theta y_{i,t-1} + x_{it}'\beta + \varepsilon_{it}, \qquad i = 1, \ldots, N, \quad t = 1, \ldots, T, \tag{34}$$

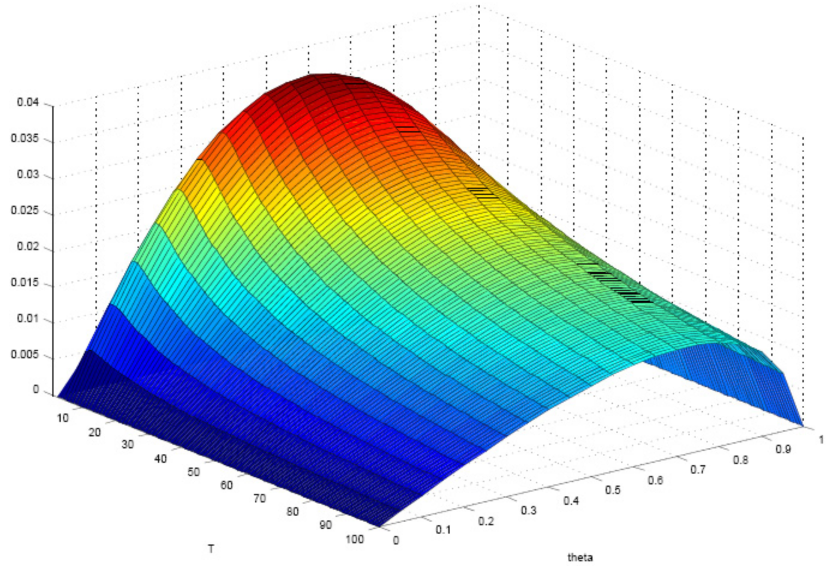
where $x_{it} = (x_{1it}, \ldots, x_{Kit})'$ is a $(K \times 1)$ vector of explanatory variables for unit i at time period t. We further assume:

Assumption A4.

$$E[x_{it}\varepsilon_{js}] = 0, \qquad \forall i, j, t, s, \tag{35}$$

$$E[x_{it}\alpha_i] = \sigma_{x\alpha}, \tag{36}$$

Figure 1: Upper bound $(\sigma_\alpha^2/\sigma_\varepsilon^2 \to \infty)$ of the inconsistency of $\widehat{\theta}_\perp^{WG}$ for $N \to \infty$



where $\sigma_{x\alpha}$ is a $K \times 1$ vector.

Assumption A4 states that the variables in x_{it} are strictly exogenous with respect to ε_{it} but allowed to be correlated with the individual effect α_i.

The most straightforward way to extend the WGob estimator defined in (18) to the case of exogenous variables is by using the Hausman and Taylor (1981) approach. By virtue of strict exogeneity of x_{it} we have, next to (19), the following set of moment conditions available

$$E\left(\widetilde{x}_{it}\left(\alpha_i + \varepsilon_{it}\right)\right) = 0, \tag{37}$$

where $\widetilde{x}_{it}$ is x_{it} in deviation from its individual-specific sample mean.

Remark 3. The strong advantage of the Hausman-Taylor approach is that the transformation of the explanatory variables required to wipe out the correlation with the composite error term $\alpha_i + \varepsilon_{it}$ can be adjusted in terms of the alleged dependency on the individual effect, the idiosyncratic error term or both. Moreover, next to internal instruments, when available, also external instruments can be added. In this paper we focus on the standard case of strictly exogenous variables which are correlated with the individual effects such that a within transformation is optimal.

Using the moment conditions in (19) and (37), the IV estimator for $\gamma = [\theta, \beta']'$ in (34) is given by

$$\widehat{\gamma}_\perp^{WG} = \left(Z'W\right)^{-1} Z'y, \tag{38}$$

where $W = [y_{-1}, X]$ and $Z = \left[Q_{\perp}^{b} y_{-1}, QX\right]$. Inserting (34) in (38) and letting $N \to \infty$, we have

$$\underset{N\to\infty}{\text{plim}} \left(\widehat{\gamma} - \gamma\right) = \underset{N\to\infty}{\text{plim}} \left(\frac{1}{NT} Z'W\right)^{-1} \underset{N\to\infty}{\text{plim}} \frac{1}{NT} Z' \left(D\alpha + \varepsilon\right). \tag{39}$$

First, by virtue of $Q_{\perp}^{b} y_{-1}$ being predetermined and QX being exogenous we have

$$\underset{N\to\infty}{\text{plim}} \frac{1}{NT} Z' \left(D\alpha + \varepsilon\right) = \begin{bmatrix} \underset{N\to\infty}{\text{plim}} \frac{1}{NT} y_{-1}' Q_{\perp}^{b} D\alpha \\ 0 \end{bmatrix} = \begin{bmatrix} \left(1 - \delta_T\right) \frac{\sigma_\alpha^2 + \beta' \sigma_{x\alpha}}{1-\theta} \\ 0 \end{bmatrix}.$$

where $\delta_T = \underset{N\to\infty}{\text{plim}} \widehat{\delta}_T$ with $\widehat{\delta}_T$ as defined in (22).

Second,

$$\left(\frac{Z'W}{NT}\right)^{-1} = \begin{bmatrix} 0 & 0 \\ 0 & \left(\frac{X'QX}{NT}\right)^{-1} \end{bmatrix} + \frac{1}{\phi} \begin{bmatrix} -1 \\ \left(X'QX\right)^{-1} X'Qy_{-1} \end{bmatrix} \begin{bmatrix} -1 & \left(X'QX\right)^{-1} y_{-1}' Q_{\perp}^{b} X \end{bmatrix},$$

where

$$\phi = \frac{1}{NT} \left(y_{-1}' Q_{\perp}^{b} y_{-1} - y_{-1}' Q_{\perp}^{b} X \left(X'QX\right)^{-1} X'Qy_{-1}\right) = \frac{1}{NT} \left(y_{-1}' \widetilde{Q}_{\perp}^{b} y_{-1}\right),$$

with $\widetilde{Q}_{\perp}^{b} = Q_{\perp}^{b} \left(I - X \left(X'QX\right)^{-1} X'Q\right)$. Writing (34) in component form

$$y_{it} = y_{it}^{0} + x_{it}^{0\prime} \beta, \quad y_{it}^{0} = \alpha_i + \theta y_{i,t-1}^{0} + \varepsilon_{it}, \quad x_{it}^{0} = \left(1 - \theta L\right)^{-1} x_{it}, \tag{40}$$

with its lagged variant using stacked notation being $y_{-1} = y_{-1}^{0} + X_{-1}^{0} \beta$, we have

$$\underset{N\to\infty}{\text{plim}} \frac{1}{NT} \left(y_{-1}' \widetilde{Q}_{\perp}^{b} y_{-1}\right) = \underset{N\to\infty}{\text{plim}} \frac{1}{NT} \left(y_{-1}^{0\prime} Q_{\perp}^{b} y_{-1}^{0}\right) + \beta' \underset{N\to\infty}{\text{plim}} \frac{1}{NT} \left(X_{-1}^{0\prime} \widetilde{Q}_{\perp}^{b} X_{-1}^{0}\right) \beta,$$

such that (39) can be rewritten as

$$\underset{N\to\infty}{\text{plim}} \left(\widehat{\theta}_{\perp}^{WG} - \theta\right) = \frac{\left(1 - \delta_T\right)}{\sigma_{\breve{y}^{0b},T}^2 + \beta' \underset{N\to\infty}{\text{plim}} \frac{1}{NT} \left(X_{-1}^{0\prime} \widetilde{Q}_{\perp}^{b} X_{-1}^{0}\right) \beta} \frac{\sigma_\alpha^2 + \beta' \sigma_{x\alpha}}{1 - \theta}, \tag{41}$$

$$\underset{N\to\infty}{\text{plim}} \left(\widehat{\beta}_{\perp}^{WG} - \beta\right) = -\left(\underset{N\to\infty}{\text{plim}} \left(X'QX\right)^{-1} X'Qy_{-1}\right) \underset{N\to\infty}{\text{plim}} \left(\widehat{\theta}_{\perp}^{WG} - \theta\right), \tag{42}$$

where $\sigma_{\breve{y}^{0b},T}^2 = \underset{N\to\infty}{\text{plim}} \frac{1}{NT} \left(y_{-1}^{0\prime} Q_{\perp}^{b} y_{-1}^{0}\right)$.

Note that the inconsistency of $\widehat{\theta}_{\perp}^{WG}$ is highly similar to the inconsistency in the case of a model with no exogenous variables. When $\beta = 0$, the inconsistency (41) collapses to (23). When $\beta \neq 0$, the relative importance of the individual effects and the idiosyncratic errors in x_{it} have a similar impact on δ_T as σ_α^2 and σ_ε^2 respectively. Two channels through which the inclusion of exogenous variables tend to reduce the inconsistency are that (i) the individual effects in x_{it} drive δ_T towards

1 but only drive up the correlation between $\breve{y}_{it}^b$ and α_i to the degree that $\sigma_{x\alpha}$ differs from zero and (ii) the denominator of (41) is larger than in the case of no exogenous variables.

6 Monte Carlo study of finite sample properties

6.1 Design of data generating process

Data are generated from (34) under a number of additional assumptions. First, we consider a single explanatory variable x_{it} which is generated as

$$x_{it} = \omega_i + \rho x_{i,t-1} + \xi_{it}, \tag{43}$$

$$\omega_i = \eta_i + \gamma\alpha_i, \tag{44}$$

where $|\rho| < 1$, $\eta_i \sim N(0, \sigma_\eta^2)$, $\alpha_i \sim N\left(0, \sigma_\alpha^2\right)$ and $\xi_{it} \sim N(0, \sigma_\xi^2)$. For, $\gamma \neq 0$ the individual effect in x_{it} is correlated with the individual effect in y_{it}. The initial value x_{i0} is drawn from the stationary distribution of x_{it} which has mean $(\eta_i + \gamma\alpha_i)(1-\rho)^{-1}$ and variance $\sigma_\xi^2\left(1-\rho^2\right)^{-1}$. This gives

$$x_{i0} = \frac{\eta_i + \gamma\alpha_i}{1-\rho} + \xi_{i0}\left(\frac{1}{1-\rho^2}\right)^{1/2}.$$

Second, y_{it} is generated from (34) using the data for x_{it} and making the additional distributional assumption $\varepsilon_{it} \sim N\left(0, \sigma_\varepsilon^2\right)$. The initial value y_{i0} is drawn from its stationary distribution which has mean $\frac{\alpha_i + \beta\omega_i(1-\rho)^{-1}}{1-\theta}$ and variance $\frac{\beta^2\sigma_\xi^2(1+\theta\rho)}{(1-\theta\rho)(1-\theta^2)(1-\rho^2)} + \frac{\sigma_\varepsilon^2}{1-\theta^2}$. This gives

$$y_{i0} = \frac{\alpha_i + \beta\omega_i(1-\rho)^{-1}}{1-\theta} + \beta\xi_{i0}\left(\frac{(1+\theta\rho)}{(1-\theta\rho)\left(1-\theta^2\right)\left(1-\rho^2\right)}\right)^{1/2} + \varepsilon_{i0}\left(\frac{1}{1-\theta^2}\right)^{1/2}.$$

Third, we impose

$$\begin{aligned}
\sigma_\varepsilon^2 &= 1, \\
\beta &= 1-\theta, \\
\sigma_\alpha^2 &= \mu_\alpha\sigma_\varepsilon^2(1-\theta)^2, \quad \text{with} \quad \mu_\alpha > 0, \\
\sigma_\omega^2 &= \mu_\omega\sigma_\xi^2(1-\rho)^2, \quad \text{with} \quad \mu_\omega > 0, \\
\sigma_\eta^2 &= \sigma_\omega^2 - \gamma^2\sigma_\alpha^2, \qquad \text{with} \quad \gamma^2 \leq \sigma_\omega^2/\sigma_\alpha^2,
\end{aligned}$$

where $\sigma_\omega^2 = V(\omega_i)$. The first two are normalizing restrictions, setting the variance of ε_{it} and the long-run impact of x_{it} on y_{it} to unity. The next two restrictions allow to control, through the values of μ_α and μ_ω, the relative impact of the disturbances ε_{it} and ξ_{it} versus the individual effects α_i and ω_i on y_{it} and x_{it} respectively. For a given value of σ_ω^2, the last restriction adjusts σ_η^2 over alternative values for γ and σ_α^2.

Finally, we calculate σ_s^2

$$\sigma_s^2 = \text{var}\left[y_{it} - \left(\frac{\alpha_i + \beta\omega_i (1-\rho)^{-1}}{1-\theta} + \varepsilon_{it}\right)\right],$$
$$= \beta^2 \frac{1+\theta\rho}{(1-\theta\rho)(1-\theta^2)(1-\rho^2)}\sigma_\xi^2 + \frac{\theta^2}{1-\theta^2}\sigma_\varepsilon^2, \tag{45}$$

which measures the variance of the signal with respect to explaining y_{it} contained in the within variation of x_{it} and y_{it-1} relative to the noise contained in α_i, ω_i and ε_{it}. Note that σ_s^2 varies with, among others, θ and ρ. As Kiviet (1995) argues that varying σ_s^2 may significantly alter the relative bias of the various estimators, we generate data controlling σ_s^2 by fixing it at some value and adjusting σ_ξ^2 over alternative values of θ and ρ.[2]

We performed 5000 Monte Carlo replications for each of the following experiments: $\theta \in \{0.4, 0.8\}$, $\rho \in \{0.4, 0.8\}$, $\mu_\alpha \in \{1, 5\}$, $\mu_\omega \in \{1, 5\}$, $\sigma_s^2 \in \{2, 8\}$, $\gamma \in \{0, 1\}$, $(T, N) \in \{(5, 20),$ $(10, 20)$, $(20, 20)$, $(5, 100)$, $(10, 100)$, $(5, 500)\}$. A selection of the results is reported in the next section.

6.2 Estimators

The performance of the WGob estimator is compared to 4 alternative dynamic panel data estimators: (i) WG, the standard within groups estimator, (ii) WGbc, the bias-corrected WG estimator proposed by Kiviet (1995), (iii) GMMd, the first difference GMM estimator proposed by Arellano and Bond (1991) and (iv) GMMs, the system GMM estimator proposed by Arellano and Bover (1995) and Blundell and Bond (1998). We opt for Kiviet's bias-corrected WG estimator over alternative, more generally applicable, bias-corrections proposed by e.g. Bun and Carree (2005) and Everaert and Pozzi (2007) as the conditions under which it is derived are satisfied in the proposed Monte Carlo design. To implement the WGbc estimator we use the GMMs estimator as an initial large-N consistent estimator. For both GMM estimators we report second-step estimates. In order to avoid an overfitting bias (see Ziliak, 1997) we restrict the number of lagged instruments to a maximum of 3 and stack instruments when $T \geq 10$ (see also Arellano, 2003, p. 170).

The estimators are compared in terms of (i) mean bias (bias), (ii) mean standard deviation (stdv), (iii) mean estimated standard deviation (stde), (iv) root mean squared error (rmse), (v) size and (vi) power. For both the WG and the WGbc estimator, stde is obtained from the standard covariance matrix $\widehat{\sigma}_\varepsilon^2 (W'QW)^{-1}$ while for the GMM estimators the corrected second-step covariance matrix (Windmeijer, 2005; Bond and Windmeijer, 2005) is used. As the WGob estimator leaves a particular autocorrelation structure in the error terms, we use the feasible robust

[2] Note that some parameter configurations are infeasible as they imply $\sigma_\xi^2 < 0$.

covariance matrix given by

$$\widehat{Var}\left(\widehat{\gamma}_{\perp}^{WG}\right) = \left(W'Z\left(\widehat{Z'\eta\eta'Z}\right)^{-1}Z'W\right)^{-1},$$
$$= \left(W'Z\left(\sum_{i=1}^{N}\begin{bmatrix}\breve{y}_{i,-1}^{b\prime}\widehat{\breve{\eta}}_i\\ \widetilde{X}_i'\widehat{\widetilde{\eta}}_i\end{bmatrix}\begin{bmatrix}\widehat{\breve{\eta}}_i'\breve{y}_{i,-1}^{b} & \widehat{\widetilde{\eta}}_i'\widetilde{X}_i\end{bmatrix}\right)^{-1}Z'W\right)^{-1}, \quad (46)$$

where $\widehat{\breve{\eta}} = Q_{\perp}^{b}\widehat{\eta}$, $\widehat{\widetilde{\eta}} = Q\widehat{\eta}$ and $\widehat{\eta} = y - W\widehat{\gamma}_{\perp}^{WG}$.

6.3 Simulation results

6.3.1 Estimating θ and β

The results of the Monte Carlo experiments in terms of estimation are presented in Table 1. A number of conclusions known from the existing literature can immediately be drawn: (i) the WG estimator for θ is severely biased, especially when θ approaches 1, with the bias disappearing as T grows large but not as N increases (see Nickell, 1981); (ii) the WG estimator for β is much less biased (Kiviet, 1995); (iii) in samples with limited N and T the GMMd estimator performs poorly both in terms of bias and dispersion when θ approaches 1 and/or when μ_α becomes large (see Blundell and Bond, 1998); (iv) the GMMs estimator improves significantly on the performance of the GMMd estimator (see Blundell and Bond, 1998) but remains biased in samples with limited N and T especially when θ is small and μ_α is large (see Kiviet, 2006); (v) the WGbc estimator outperforms the GMMs estimator in a lot of cases as it successfully succeeds in removing the bias from the WG estimator while maintaining its relatively small dispersion, however, it remains biased when T is small especially when both θ and μ_α are large (Kiviet, 1995; Judson and Owen, 1999).

Turning to the WGob estimator, it performs remarkably well in terms of bias for both θ and β, i.e. despite a slight tendency to increase in γ and μ_α the bias is negligibly small in all of the experiments. As a result the WGob estimator clearly outperforms the other estimators in terms of bias in the cases where these estimators are (severely) biased while not being much worse in the cases where these estimators are unbiased/consistent. The dispersion, measured by the stdv, of the WGob is larger than that of the WG and WGbc estimators, especially when T is small, but smaller than that of the GMM estimators in most cases. This implies that in terms of rmse, the WGob estimator is slightly outperformed by the WGbc estimator in a number of cases, while being significantly smaller in others, but outperforms the GMMs estimator in almost all cases, interestingly even when $N = 500$.

Table 1: Monte Carlo results

			θ = 0.4								θ = 0.8							
			Results for θ				Results for β				Results for θ				Results for β			
T	N		bias	stde	stdv	rmse	bias	stde	stdv	rmse	bias	stde	stdv	rmse	bias	stde	stdv	rmse
Experiment 1: $\rho = 0.4,\ \sigma_s^2 = 2.0,\ \mu_\alpha = 1.0,\ \mu_\omega = 1.0,\ \gamma = 0.0$																		
5	20	WG	−0.149	0.080	0.080	0.169	0.002	0.072	0.074	0.074	−0.419	0.103	0.110	0.433	−0.016	0.116	0.122	0.123
		WGbc	0.005	0.082	0.086	0.086	0.004	0.074	0.073	0.073	−0.106	0.110	0.108	0.151	−0.002	0.123	0.124	0.124
		WGob	0.003	0.111	0.120	0.120	0.003	0.070	0.074	0.074	−0.016	0.131	0.143	0.144	0.004	0.121	0.128	0.128
		GMMd	−0.063	0.129	0.132	0.147	−0.001	0.098	0.101	0.101	−0.382	0.267	0.281	0.474	−0.016	0.160	0.164	0.165
		GMMs	0.042	0.112	0.100	0.109	−0.010	0.096	0.089	0.090	−0.009	0.115	0.123	0.123	0.005	0.140	0.143	0.143
10	20	WG	−0.068	0.049	0.050	0.085	0.008	0.046	0.047	0.047	−0.212	0.059	0.065	0.222	−0.002	0.077	0.081	0.081
		WGbc	0.001	0.049	0.053	0.053	−0.001	0.046	0.047	0.047	−0.032	0.061	0.075	0.081	−0.001	0.079	0.081	0.081
		WGob	0.007	0.065	0.071	0.071	−0.002	0.046	0.048	0.048	−0.005	0.065	0.071	0.071	−0.001	0.076	0.079	0.079
		GMMd	−0.002	0.077	0.080	0.080	−0.004	0.062	0.064	0.064	−0.032	0.166	0.170	0.173	−0.004	0.109	0.113	0.113
		GMMs	0.005	0.073	0.076	0.076	−0.005	0.061	0.063	0.064	−0.009	0.116	0.122	0.123	−0.004	0.110	0.114	0.114
20	20	WG	−0.033	0.032	0.032	0.046	0.006	0.031	0.031	0.032	−0.102	0.036	0.039	0.109	0.003	0.052	0.055	0.055
		WGbc	0.000	0.032	0.033	0.033	−0.001	0.031	0.031	0.031	−0.006	0.036	0.045	0.045	0.000	0.053	0.054	0.054
		WGob	0.009	0.042	0.045	0.046	−0.002	0.031	0.033	0.033	0.002	0.037	0.041	0.041	−0.001	0.051	0.053	0.053
		GMMd	0.000	0.047	0.049	0.049	−0.003	0.040	0.042	0.042	−0.004	0.086	0.090	0.090	−0.002	0.072	0.076	0.076
		GMMs	0.003	0.046	0.048	0.048	−0.003	0.041	0.042	0.042	0.000	0.071	0.074	0.074	−0.002	0.073	0.077	0.077
5	100	WG	−0.146	0.035	0.036	0.151	0.000	0.032	0.033	0.033	−0.410	0.046	0.050	0.413	−0.018	0.051	0.054	0.057
		WGbc	−0.005	0.036	0.041	0.041	0.001	0.033	0.032	0.032	−0.092	0.048	0.058	0.109	−0.004	0.054	0.056	0.056
		WGob	0.007	0.052	0.054	0.054	0.000	0.032	0.033	0.033	−0.001	0.060	0.063	0.063	0.001	0.055	0.056	0.056
		GMMd	−0.012	0.060	0.061	0.062	−0.001	0.044	0.045	0.045	−0.112	0.158	0.158	0.194	−0.004	0.077	0.079	0.079
		GMMs	0.008	0.053	0.055	0.055	−0.003	0.043	0.044	0.044	−0.003	0.083	0.086	0.086	0.000	0.076	0.079	0.078
10	100	WG	−0.067	0.022	0.021	0.070	0.009	0.020	0.020	0.022	−0.207	0.026	0.029	0.209	−0.001	0.034	0.035	0.035
		WGbc	−0.001	0.022	0.023	0.023	0.000	0.021	0.020	0.020	−0.026	0.027	0.035	0.044	0.000	0.035	0.035	0.035
		WGob	0.010	0.031	0.031	0.033	−0.001	0.021	0.021	0.021	0.005	0.030	0.030	0.031	0.000	0.035	0.034	0.034
		GMMd	−0.001	0.033	0.034	0.034	0.000	0.027	0.027	0.027	−0.006	0.070	0.071	0.071	0.000	0.048	0.048	0.048
		GMMs	0.001	0.032	0.033	0.033	−0.001	0.027	0.026	0.026	−0.003	0.053	0.054	0.054	0.000	0.048	0.048	0.048
5	500	WG	−0.144	0.016	0.016	0.145	−0.001	0.014	0.015	0.015	−0.407	0.020	0.023	0.408	−0.019	0.023	0.024	0.031
		WGbc	−0.005	0.016	0.018	0.019	0.000	0.015	0.015	0.015	−0.087	0.022	0.028	0.091	−0.004	0.024	0.025	0.026
		WGob	0.010	0.023	0.024	0.027	0.000	0.015	0.015	0.015	0.005	0.027	0.028	0.028	0.000	0.025	0.025	0.025
		GMMd	−0.001	0.026	0.026	0.026	0.001	0.019	0.019	0.019	−0.020	0.069	0.067	0.070	0.001	0.034	0.034	0.034
		GMMs	0.002	0.023	0.023	0.023	0.000	0.018	0.018	0.018	0.001	0.037	0.038	0.038	0.001	0.033	0.033	0.033

Continued on next page

Continued from previous page

			$\theta = 0.4$								$\theta = 0.8$							
			Results for θ				Results for β				Results for θ				Results for β			
T	N		bias	stde	stdv	rmse	bias	stde	stdv	rmse	bias	stde	stdv	rmse	bias	stde	stdv	rmse
Experiment 2: $\rho = 0.4,\ \sigma_s^2 = 2.0,\ \mu_\alpha = 1.0,\ \mu_\omega = 1.0,\ \gamma = 1.0$																		
5	20	WG	−0.149	0.080	0.080	0.169	0.002	0.072	0.074	0.074	−0.419	0.103	0.110	0.433	−0.016	0.116	0.122	0.123
		WGbc	0.006	0.082	0.085	0.085	0.005	0.074	0.073	0.073	−0.105	0.110	0.107	0.150	−0.001	0.123	0.124	0.124
		WGob	0.006	0.103	0.111	0.111	0.003	0.070	0.074	0.074	−0.016	0.130	0.142	0.143	0.004	0.121	0.128	0.128
		GMMd	−0.060	0.126	0.129	0.142	0.001	0.097	0.100	0.100	−0.376	0.265	0.279	0.469	−0.013	0.160	0.164	0.165
		GMMs	0.067	0.087	0.093	0.114	0.021	0.084	0.088	0.090	−0.002	0.111	0.118	0.118	0.018	0.142	0.145	0.146
10	20	WG	−0.068	0.049	0.050	0.085	0.008	0.046	0.047	0.047	−0.212	0.059	0.065	0.222	−0.002	0.077	0.081	0.081
		WGbc	0.001	0.049	0.053	0.053	−0.001	0.046	0.046	0.046	−0.032	0.061	0.075	0.081	−0.001	0.079	0.081	0.081
		WGob	0.012	0.059	0.064	0.065	−0.003	0.045	0.048	0.048	−0.004	0.065	0.071	0.071	−0.001	0.076	0.079	0.079
		GMMd	−0.002	0.077	0.079	0.079	−0.004	0.062	0.064	0.064	−0.033	0.167	0.171	0.174	−0.003	0.109	0.113	0.113
		GMMs	0.009	0.073	0.077	0.077	−0.001	0.062	0.063	0.063	−0.008	0.116	0.123	0.123	−0.003	0.110	0.114	0.114
20	20	WG	−0.033	0.032	0.032	0.046	0.006	0.031	0.031	0.032	−0.102	0.036	0.039	0.109	0.003	0.052	0.055	0.055
		WGbc	0.000	0.032	0.033	0.033	−0.001	0.031	0.031	0.031	−0.006	0.036	0.045	0.045	0.000	0.053	0.054	0.054
		WGob	0.012	0.037	0.040	0.042	−0.003	0.031	0.032	0.032	0.003	0.037	0.040	0.040	−0.001	0.051	0.053	0.053
		GMMd	0.001	0.047	0.049	0.049	−0.002	0.041	0.042	0.042	−0.004	0.088	0.092	0.092	−0.002	0.072	0.076	0.076
		GMMs	0.005	0.047	0.049	0.049	−0.001	0.041	0.042	0.042	0.001	0.072	0.074	0.074	−0.002	0.073	0.077	0.077
5	100	WG	−0.146	0.035	0.036	0.151	0.000	0.032	0.033	0.033	−0.410	0.046	0.050	0.413	−0.018	0.051	0.054	0.057
		WGbc	−0.004	0.036	0.041	0.041	0.001	0.033	0.032	0.032	−0.091	0.048	0.058	0.108	−0.003	0.054	0.056	0.056
		WGob	0.011	0.048	0.050	0.051	0.000	0.032	0.033	0.033	0.000	0.059	0.062	0.062	0.001	0.055	0.056	0.056
		GMMd	−0.012	0.058	0.060	0.061	−0.001	0.044	0.044	0.044	−0.110	0.156	0.157	0.192	−0.004	0.076	0.078	0.078
		GMMs	0.017	0.053	0.055	0.057	0.004	0.043	0.044	0.044	0.000	0.082	0.085	0.085	0.003	0.077	0.079	0.079
10	100	WG	−0.067	0.022	0.021	0.070	0.009	0.020	0.020	0.022	−0.207	0.026	0.029	0.209	−0.001	0.034	0.035	0.035
		WGbc	−0.001	0.022	0.023	0.023	0.000	0.021	0.020	0.020	−0.026	0.027	0.035	0.044	0.000	0.035	0.035	0.035
		WGob	0.015	0.028	0.028	0.032	−0.002	0.021	0.021	0.021	0.005	0.030	0.030	0.031	0.000	0.035	0.034	0.034
		GMMd	−0.001	0.033	0.034	0.034	0.000	0.027	0.027	0.027	−0.006	0.071	0.072	0.073	0.000	0.048	0.048	0.048
		GMMs	0.002	0.032	0.033	0.033	0.000	0.027	0.026	0.026	−0.003	0.054	0.054	0.054	0.000	0.048	0.048	0.048
5	500	WG	−0.144	0.016	0.016	0.145	−0.001	0.014	0.015	0.015	−0.407	0.020	0.023	0.408	−0.019	0.023	0.024	0.031
		WGbc	−0.005	0.016	0.018	0.019	0.000	0.015	0.015	0.015	−0.087	0.022	0.028	0.091	−0.004	0.024	0.025	0.025
		WGob	0.015	0.022	0.022	0.027	0.000	0.015	0.015	0.015	0.006	0.027	0.028	0.028	0.000	0.025	0.025	0.025
		GMMd	−0.001	0.025	0.025	0.025	0.000	0.019	0.019	0.019	−0.019	0.068	0.066	0.069	0.001	0.034	0.034	0.034
		GMMs	0.004	0.023	0.023	0.024	0.001	0.018	0.018	0.018	0.002	0.037	0.038	0.038	0.001	0.033	0.033	0.033

Continued on next page

Continued from previous page

			θ = 0.4								θ = 0.8							
			Results for θ				Results for β				Results for θ				Results for β			
T	N		bias	stde	stdv	rmse	bias	stde	stdv	rmse	bias	stde	stdv	rmse	bias	stde	stdv	rmse
Experiment 3: $\rho = 0.8,\ \sigma_s^2 = 2.0,\ \mu_\alpha = 1.0,\ \mu_\omega = 1.0,\ \gamma = 1.0$																		
5	20	WG	−0.225	0.097	0.095	0.244	0.051	0.129	0.136	0.146	−0.435	0.105	0.112	0.449	0.024	0.267	0.299	0.300
		WGbc	−0.016	0.100	0.107	0.108	0.010	0.133	0.136	0.136	−0.118	0.111	0.111	0.162	0.020	0.282	0.285	0.285
		WGob	−0.003	0.129	0.139	0.139	0.004	0.137	0.146	0.147	−0.019	0.141	0.153	0.154	0.009	0.283	0.303	0.303
		GMMd	−0.104	0.161	0.165	0.195	0.027	0.185	0.189	0.191	−0.386	0.266	0.287	0.481	0.033	0.393	0.403	0.404
		GMMs	0.039	0.106	0.110	0.116	0.089	0.139	0.143	0.168	−0.024	0.122	0.128	0.130	0.131	0.271	0.278	0.307
10	20	WG	−0.102	0.059	0.060	0.119	0.041	0.075	0.078	0.089	−0.221	0.060	0.066	0.231	0.026	0.155	0.178	0.180
		WGbc	−0.001	0.060	0.067	0.067	−0.001	0.076	0.078	0.078	−0.037	0.062	0.076	0.084	0.002	0.159	0.171	0.171
		WGob	0.012	0.072	0.079	0.080	−0.007	0.081	0.086	0.086	−0.005	0.068	0.073	0.073	−0.003	0.155	0.164	0.164
		GMMd	−0.003	0.102	0.107	0.107	−0.004	0.122	0.126	0.126	−0.033	0.168	0.173	0.176	−0.002	0.271	0.279	0.279
		GMMs	0.017	0.095	0.100	0.101	0.015	0.120	0.122	0.123	−0.009	0.122	0.128	0.129	0.012	0.263	0.270	0.270
20	20	WG	−0.048	0.039	0.039	0.062	0.026	0.048	0.049	0.056	−0.107	0.036	0.039	0.114	0.024	0.094	0.108	0.110
		WGbc	0.000	0.039	0.040	0.040	−0.001	0.048	0.049	0.049	−0.007	0.037	0.045	0.046	0.001	0.095	0.102	0.102
		WGob	0.017	0.045	0.047	0.050	−0.010	0.050	0.054	0.055	0.003	0.038	0.040	0.040	−0.002	0.092	0.099	0.099
		GMMd	0.001	0.063	0.066	0.066	−0.003	0.079	0.083	0.083	−0.004	0.089	0.093	0.093	−0.003	0.176	0.185	0.185
		GMMs	0.009	0.062	0.064	0.065	0.007	0.079	0.081	0.082	0.001	0.075	0.078	0.078	0.004	0.174	0.181	0.181
5	100	WG	−0.218	0.043	0.043	0.223	0.047	0.057	0.060	0.076	−0.424	0.046	0.050	0.427	0.017	0.117	0.132	0.133
		WGbc	−0.012	0.044	0.052	0.053	0.004	0.059	0.060	0.060	−0.098	0.049	0.058	0.114	0.007	0.125	0.129	0.129
		WGob	0.009	0.059	0.061	0.062	−0.002	0.062	0.064	0.064	0.000	0.061	0.064	0.064	0.000	0.129	0.131	0.131
		GMMd	−0.021	0.077	0.078	0.081	0.005	0.084	0.085	0.085	−0.112	0.158	0.158	0.194	0.011	0.187	0.192	0.192
		GMMs	0.026	0.064	0.066	0.071	0.037	0.080	0.082	0.090	−0.002	0.084	0.086	0.086	0.042	0.175	0.180	0.185
10	100	WG	−0.099	0.026	0.026	0.102	0.041	0.033	0.034	0.053	−0.214	0.027	0.029	0.216	0.028	0.069	0.077	0.082
		WGbc	−0.002	0.027	0.029	0.029	0.000	0.034	0.034	0.034	−0.028	0.028	0.036	0.046	0.004	0.070	0.073	0.073
		WGob	0.017	0.034	0.034	0.038	−0.008	0.037	0.037	0.038	0.007	0.031	0.031	0.032	−0.002	0.071	0.071	0.071
		GMMd	−0.001	0.045	0.045	0.045	0.001	0.054	0.053	0.053	−0.006	0.073	0.073	0.073	0.001	0.119	0.120	0.120
		GMMs	0.004	0.043	0.043	0.044	0.006	0.053	0.052	0.053	−0.002	0.056	0.057	0.057	0.005	0.118	0.120	0.120
5	500	WG	−0.215	0.019	0.019	0.216	0.046	0.025	0.027	0.053	−0.422	0.021	0.022	0.422	0.016	0.052	0.059	0.061
		WGbc	−0.012	0.020	0.024	0.026	0.003	0.026	0.027	0.027	−0.093	0.022	0.027	0.097	0.004	0.056	0.058	0.059
		WGob	0.014	0.026	0.027	0.030	−0.003	0.028	0.028	0.028	0.006	0.027	0.028	0.028	0.000	0.058	0.058	0.058
		GMMd	−0.003	0.033	0.033	0.033	0.002	0.036	0.036	0.036	−0.020	0.068	0.067	0.070	0.002	0.082	0.083	0.083
		GMMs	0.008	0.029	0.030	0.031	0.010	0.036	0.035	0.037	0.001	0.038	0.038	0.038	0.010	0.081	0.081	0.082

Continued on next page

Continued from previous page

			θ = 0.4								θ = 0.8							
			Results for θ				Results for β				Results for θ				Results for β			
T	N		bias	stde	stdv	rmse	bias	stde	stdv	rmse	bias	stde	stdv	rmse	bias	stde	stdv	rmse
Experiment 4: $\rho = 0.4,\ \sigma_s^2 = 2.0,\ \mu_\alpha = 5.0,\ \mu_\omega = 1.0,\ \gamma = 1.0$																		
5	20	WG	−0.149	0.080	0.080	0.169	0.002	0.072	0.074	0.074	−0.419	0.103	0.110	0.433	−0.016	0.116	0.122	0.123
		WGbc	0.028	0.083	0.081	0.086	0.011	0.074	0.073	0.074	−0.088	0.110	0.099	0.132	0.004	0.123	0.125	0.125
		WGob	0.013	0.099	0.107	0.108	0.002	0.070	0.074	0.074	−0.012	0.129	0.142	0.142	0.004	0.121	0.128	0.128
		GMMd	−0.060	0.127	0.130	0.143	0.000	0.097	0.100	0.100	−0.417	0.276	0.295	0.511	−0.015	0.158	0.162	0.163
		GMMs	0.176	0.073	0.080	0.193	0.059	0.089	0.094	0.111	0.065	0.089	0.093	0.113	0.035	0.148	0.151	0.155
10	20	WG	−0.068	0.049	0.050	0.085	0.008	0.046	0.047	0.047	−0.212	0.059	0.065	0.222	−0.002	0.077	0.081	0.081
		WGbc	0.001	0.049	0.053	0.053	0.000	0.046	0.046	0.046	−0.030	0.061	0.074	0.080	0.000	0.079	0.081	0.081
		WGob	0.022	0.056	0.061	0.064	−0.004	0.045	0.048	0.048	0.003	0.064	0.070	0.070	−0.001	0.076	0.079	0.079
		GMMd	−0.003	0.080	0.082	0.082	−0.003	0.062	0.064	0.065	−0.048	0.195	0.200	0.206	−0.002	0.109	0.112	0.112
		GMMs	0.032	0.079	0.086	0.091	0.007	0.064	0.066	0.066	0.006	0.123	0.131	0.131	0.000	0.112	0.115	0.115
20	20	WG	−0.033	0.032	0.032	0.046	0.006	0.031	0.031	0.032	−0.102	0.036	0.039	0.109	0.003	0.052	0.055	0.055
		WGbc	0.000	0.032	0.033	0.033	−0.001	0.031	0.031	0.031	−0.005	0.036	0.045	0.045	0.000	0.053	0.054	0.054
		WGob	0.020	0.034	0.037	0.042	−0.005	0.030	0.032	0.032	0.011	0.036	0.039	0.041	−0.001	0.050	0.053	0.053
		GMMd	0.000	0.049	0.051	0.051	−0.002	0.041	0.042	0.043	−0.007	0.104	0.109	0.109	−0.001	0.072	0.075	0.075
		GMMs	0.014	0.050	0.053	0.055	0.002	0.042	0.043	0.043	0.008	0.077	0.080	0.081	−0.001	0.074	0.077	0.077
5	100	WG	−0.146	0.035	0.036	0.151	0.000	0.032	0.033	0.033	−0.410	0.046	0.050	0.413	−0.018	0.051	0.054	0.057
		WGbc	0.005	0.036	0.042	0.042	0.003	0.033	0.032	0.033	−0.082	0.049	0.055	0.098	−0.002	0.054	0.056	0.056
		WGob	0.020	0.046	0.048	0.052	0.000	0.032	0.033	0.033	0.006	0.059	0.062	0.063	0.001	0.056	0.056	0.056
		GMMd	−0.012	0.059	0.060	0.061	−0.001	0.044	0.044	0.044	−0.141	0.175	0.178	0.227	−0.004	0.076	0.077	0.077
		GMMs	0.069	0.060	0.068	0.097	0.017	0.046	0.046	0.049	0.034	0.080	0.083	0.090	0.008	0.079	0.080	0.081
10	100	WG	−0.067	0.022	0.021	0.070	0.009	0.020	0.020	0.022	−0.206	0.026	0.029	0.208	−0.001	0.034	0.035	0.036
		WGbc	−0.001	0.022	0.023	0.023	0.000	0.021	0.020	0.020	−0.025	0.027	0.036	0.044	0.000	0.035	0.036	0.036
		WGob	0.025	0.026	0.026	0.036	−0.003	0.021	0.021	0.021	0.014	0.030	0.030	0.033	0.000	0.035	0.035	0.035
		GMMd	−0.001	0.035	0.035	0.035	0.000	0.027	0.027	0.027	−0.008	0.085	0.085	0.086	0.001	0.048	0.048	0.048
		GMMs	0.006	0.035	0.036	0.036	0.002	0.027	0.027	0.027	0.002	0.059	0.060	0.060	0.001	0.048	0.048	0.048
5	500	WG	−0.144	0.016	0.016	0.145	−0.001	0.014	0.015	0.015	−0.408	0.020	0.022	0.409	−0.019	0.023	0.024	0.031
		WGbc	−0.004	0.016	0.019	0.019	0.000	0.015	0.015	0.015	−0.086	0.022	0.027	0.090	−0.004	0.024	0.025	0.026
		WGob	0.024	0.021	0.021	0.032	0.000	0.015	0.015	0.015	0.011	0.027	0.027	0.029	0.001	0.025	0.025	0.025
		GMMd	−0.001	0.025	0.025	0.025	0.000	0.019	0.019	0.019	−0.027	0.077	0.076	0.081	−0.001	0.033	0.034	0.034
		GMMs	0.010	0.025	0.026	0.028	0.003	0.019	0.019	0.019	0.008	0.041	0.041	0.042	0.002	0.033	0.034	0.034

Continued on next page

Continued from previous page

			$\theta = 0.4$								$\theta = 0.8$							
			Results for θ				Results for β				Results for θ				Results for β			
T	N		bias	stde	stdv	rmse	bias	stde	stdv	rmse	bias	stde	stdv	rmse	bias	stde	stdv	rmse
Experiment 5: $\rho = 0.4,\ \sigma_s^2 = 2.0,\ \mu_\alpha = 1.0,\ \mu_\omega = 5.0,\ \gamma = 1.0$																		
5	20	WG	−0.149	0.080	0.080	0.169	0.002	0.072	0.074	0.074	−0.419	0.103	0.110	0.433	−0.016	0.116	0.122	0.123
		WGbc	0.004	0.082	0.085	0.085	0.004	0.074	0.073	0.073	−0.106	0.110	0.106	0.150	−0.002	0.123	0.122	0.122
		WGob	−0.001	0.111	0.120	0.120	0.003	0.070	0.074	0.074	−0.018	0.135	0.147	0.148	0.004	0.121	0.128	0.128
		GMMd	−0.062	0.129	0.131	0.145	0.000	0.098	0.101	0.101	−0.380	0.267	0.282	0.473	−0.016	0.160	0.166	0.166
		GMMs	0.043	0.082	0.086	0.096	−0.003	0.079	0.083	0.083	−0.009	0.109	0.113	0.113	0.013	0.125	0.128	0.129
10	20	WG	−0.068	0.049	0.050	0.085	0.008	0.046	0.047	0.047	−0.212	0.059	0.065	0.222	−0.002	0.077	0.081	0.081
		WGbc	0.001	0.049	0.053	0.053	−0.001	0.046	0.046	0.046	−0.033	0.061	0.074	0.081	−0.001	0.079	0.081	0.081
		WGob	0.002	0.066	0.071	0.071	−0.001	0.046	0.048	0.048	−0.007	0.067	0.073	0.073	−0.001	0.076	0.079	0.079
		GMMd	−0.002	0.080	0.083	0.083	−0.003	0.063	0.065	0.066	−0.032	0.171	0.174	0.177	−0.003	0.111	0.115	0.115
		GMMs	0.006	0.074	0.077	0.077	−0.002	0.062	0.063	0.063	−0.011	0.120	0.126	0.126	−0.003	0.110	0.113	0.113
20	20	WG	−0.033	0.032	0.032	0.046	0.006	0.031	0.031	0.032	−0.102	0.036	0.039	0.109	0.003	0.052	0.055	0.055
		WGbc	0.000	0.032	0.033	0.033	−0.001	0.031	0.031	0.031	−0.006	0.036	0.044	0.045	0.000	0.053	0.054	0.054
		WGob	0.005	0.043	0.046	0.046	−0.002	0.031	0.033	0.033	−0.001	0.038	0.042	0.042	−0.001	0.051	0.053	0.053
		GMMd	0.000	0.049	0.051	0.051	−0.002	0.041	0.042	0.043	−0.004	0.091	0.095	0.096	−0.002	0.073	0.077	0.077
		GMMs	0.004	0.048	0.050	0.050	−0.002	0.041	0.042	0.042	0.000	0.074	0.077	0.077	−0.002	0.073	0.077	0.077
5	100	WG	−0.146	0.035	0.036	0.151	0.000	0.032	0.033	0.033	−0.410	0.046	0.050	0.413	−0.018	0.051	0.054	0.057
		WGbc	−0.004	0.036	0.040	0.041	0.001	0.033	0.032	0.032	−0.092	0.048	0.057	0.108	−0.004	0.054	0.055	0.055
		WGob	0.004	0.052	0.054	0.054	0.000	0.032	0.033	0.033	−0.002	0.060	0.063	0.063	0.000	0.055	0.056	0.056
		GMMd	−0.012	0.060	0.061	0.062	−0.001	0.044	0.045	0.045	−0.111	0.157	0.158	0.193	−0.004	0.077	0.079	0.079
		GMMs	0.014	0.050	0.052	0.054	0.001	0.042	0.043	0.043	−0.002	0.080	0.082	0.082	0.002	0.074	0.076	0.076
10	100	WG	−0.067	0.022	0.021	0.070	0.009	0.020	0.020	0.022	−0.207	0.026	0.029	0.209	−0.001	0.034	0.035	0.035
		WGbc	−0.001	0.022	0.023	0.023	0.000	0.021	0.020	0.020	−0.026	0.027	0.035	0.044	0.000	0.035	0.035	0.035
		WGob	0.006	0.031	0.031	0.032	−0.001	0.021	0.021	0.021	0.002	0.031	0.031	0.031	0.000	0.035	0.034	0.034
		GMMd	0.000	0.035	0.036	0.036	0.000	0.027	0.027	0.027	−0.006	0.074	0.074	0.075	0.000	0.049	0.049	0.049
		GMMs	0.002	0.033	0.034	0.034	0.000	0.027	0.027	0.027	−0.003	0.056	0.056	0.056	0.000	0.049	0.049	0.049
5	500	WG	−0.144	0.016	0.016	0.145	−0.001	0.014	0.015	0.015	−0.407	0.020	0.023	0.408	−0.019	0.023	0.024	0.031
		WGbc	−0.005	0.016	0.018	0.019	0.000	0.015	0.015	0.015	−0.087	0.022	0.028	0.091	−0.004	0.024	0.025	0.025
		WGob	0.007	0.023	0.025	0.025	0.000	0.014	0.015	0.015	0.004	0.027	0.028	0.028	0.000	0.025	0.025	0.025
		GMMd	−0.001	0.026	0.025	0.025	0.001	0.019	0.019	0.019	−0.020	0.068	0.067	0.070	0.001	0.034	0.034	0.034
		GMMs	0.003	0.023	0.023	0.023	0.001	0.018	0.018	0.018	0.001	0.037	0.038	0.038	0.001	0.033	0.033	0.033

Continued on next page

Continued from previous page

			$\theta = 0.4$								$\theta = 0.8$							
			Results for θ				Results for β				Results for θ				Results for β			
T	N		bias	stde	stdv	rmse	bias	stde	stdv	rmse	bias	stde	stdv	rmse	bias	stde	stdv	rmse
Experiment 6: $\rho = 0.4,\ \sigma_s^2 = 8.0,\ \mu_\alpha = 1.0,\ \mu_\omega = 1.0,\ \gamma = 1.0$																		
5	20	WG	−0.055	0.049	0.049	0.074	0.001	0.035	0.035	0.035	−0.189	0.073	0.079	0.205	−0.008	0.023	0.024	0.025
		WGbc	0.003	0.050	0.050	0.050	0.002	0.035	0.035	0.035	−0.022	0.076	0.070	0.074	0.000	0.024	0.024	0.024
		WGob	0.003	0.069	0.074	0.074	0.001	0.034	0.036	0.036	−0.009	0.086	0.094	0.094	0.001	0.022	0.023	0.023
		GMMd	−0.017	0.069	0.071	0.073	0.000	0.046	0.048	0.048	−0.075	0.126	0.130	0.150	−0.003	0.032	0.033	0.033
		GMMs	0.025	0.052	0.055	0.060	0.004	0.040	0.041	0.041	−0.001	0.050	0.053	0.053	0.002	0.026	0.027	0.027
10	20	WG	−0.024	0.029	0.029	0.038	0.003	0.022	0.022	0.023	−0.082	0.038	0.040	0.091	−0.001	0.015	0.015	0.015
		WGbc	0.000	0.029	0.030	0.030	0.000	0.022	0.022	0.022	−0.003	0.038	0.043	0.043	0.000	0.015	0.015	0.015
		WGob	0.005	0.040	0.044	0.044	−0.001	0.022	0.024	0.024	−0.003	0.042	0.045	0.046	0.000	0.014	0.015	0.015
		GMMd	−0.001	0.042	0.044	0.044	−0.002	0.029	0.030	0.031	−0.007	0.073	0.075	0.075	−0.001	0.020	0.021	0.021
		GMMs	0.003	0.041	0.042	0.042	−0.001	0.028	0.029	0.029	−0.003	0.061	0.063	0.063	−0.001	0.019	0.020	0.020
20	20	WG	−0.011	0.019	0.019	0.022	0.002	0.015	0.015	0.015	−0.037	0.022	0.022	0.043	0.001	0.010	0.010	0.010
		WGbc	0.000	0.019	0.019	0.019	0.000	0.015	0.015	0.015	0.000	0.022	0.023	0.023	0.000	0.010	0.010	0.010
		WGob	0.006	0.026	0.028	0.028	−0.002	0.015	0.016	0.016	0.000	0.024	0.026	0.026	0.000	0.010	0.010	0.010
		GMMd	0.000	0.026	0.027	0.027	−0.001	0.019	0.020	0.020	−0.001	0.043	0.045	0.045	−0.001	0.013	0.014	0.014
		GMMs	0.002	0.026	0.027	0.027	−0.001	0.019	0.019	0.019	0.000	0.040	0.042	0.042	0.000	0.013	0.013	0.013
5	100	WG	−0.054	0.022	0.022	0.058	0.000	0.016	0.016	0.016	−0.183	0.032	0.035	0.187	−0.008	0.010	0.011	0.014
		WGbc	−0.001	0.022	0.023	0.023	0.000	0.016	0.016	0.016	−0.020	0.033	0.035	0.040	−0.001	0.011	0.011	0.011
		WGob	0.005	0.032	0.033	0.033	0.000	0.016	0.016	0.016	−0.002	0.040	0.041	0.041	0.000	0.010	0.010	0.010
		GMMd	−0.003	0.031	0.031	0.032	−0.001	0.021	0.021	0.021	−0.015	0.058	0.059	0.061	−0.001	0.014	0.015	0.015
		GMMs	0.005	0.029	0.029	0.030	0.000	0.020	0.020	0.020	−0.001	0.039	0.040	0.040	0.000	0.013	0.013	0.013
10	100	WG	−0.024	0.013	0.013	0.027	0.003	0.010	0.010	0.010	−0.079	0.017	0.017	0.081	0.000	0.007	0.007	0.007
		WGbc	0.000	0.013	0.013	0.013	0.000	0.010	0.010	0.010	−0.004	0.017	0.019	0.019	0.000	0.007	0.006	0.006
		WGob	0.007	0.019	0.019	0.020	−0.001	0.010	0.010	0.010	0.000	0.019	0.020	0.020	0.000	0.007	0.006	0.006
		GMMd	0.000	0.018	0.019	0.019	0.000	0.013	0.013	0.013	−0.002	0.032	0.032	0.032	0.000	0.009	0.009	0.009
		GMMs	0.000	0.018	0.018	0.018	0.000	0.012	0.012	0.012	−0.001	0.029	0.029	0.029	0.000	0.008	0.008	0.008
5	500	WG	−0.053	0.010	0.010	0.054	0.000	0.007	0.007	0.007	−0.182	0.014	0.016	0.182	−0.008	0.005	0.005	0.010
		WGbc	0.000	0.010	0.010	0.010	0.000	0.007	0.007	0.007	−0.019	0.015	0.016	0.025	−0.001	0.005	0.005	0.005
		WGob	0.007	0.014	0.015	0.016	0.000	0.007	0.007	0.007	0.001	0.018	0.019	0.019	0.000	0.005	0.005	0.005
		GMMd	0.000	0.013	0.013	0.013	0.000	0.009	0.009	0.009	−0.003	0.025	0.024	0.024	0.000	0.006	0.006	0.006
		GMMs	0.001	0.012	0.013	0.013	0.000	0.008	0.008	0.008	0.000	0.019	0.019	0.019	0.000	0.006	0.006	0.006

6.3.2 Inference on θ

Inference is based on a standard t-test for the null hypothesis $\theta = \theta_0$ against the alternative $\theta \neq \theta_0$. In constructing the t-statistic we use the estimated standard deviations as defined in section 6.2. Table 2 contains observed sizes of a two-sided t-test with a 5% nominal size. The most important conclusion is that all estimators except WGob exhibit serious to dramatic size distortions in at least some of the experiments. For the WG estimator, the size distortion is dramatic in all experiments, even for $T = 40$ while for the WGbc estimator the size distortion is also dramatic for N growing large with T fixed (especially for $\theta = 0.8$ but disappears as T grows large. Both GMM estimators show a moderate to substantial size distortion when both T and N are small (especially when $\theta = 0.8$) for GMMd and $\mu_\alpha = 5$ for GMMs) but improve considerably when either T, N or both grow large. The WGob estimator clearly has superior size properties especially in small samples, i.e. it is more or less correctly sized when T is small with only some moderate size distortion for larger values of T when also N is large and/or $\mu_\alpha = 5$. This distortion occurs because of the relative slow disappearance of the asymptotic bias term (for N large) of the WGob estimator when T grows large.

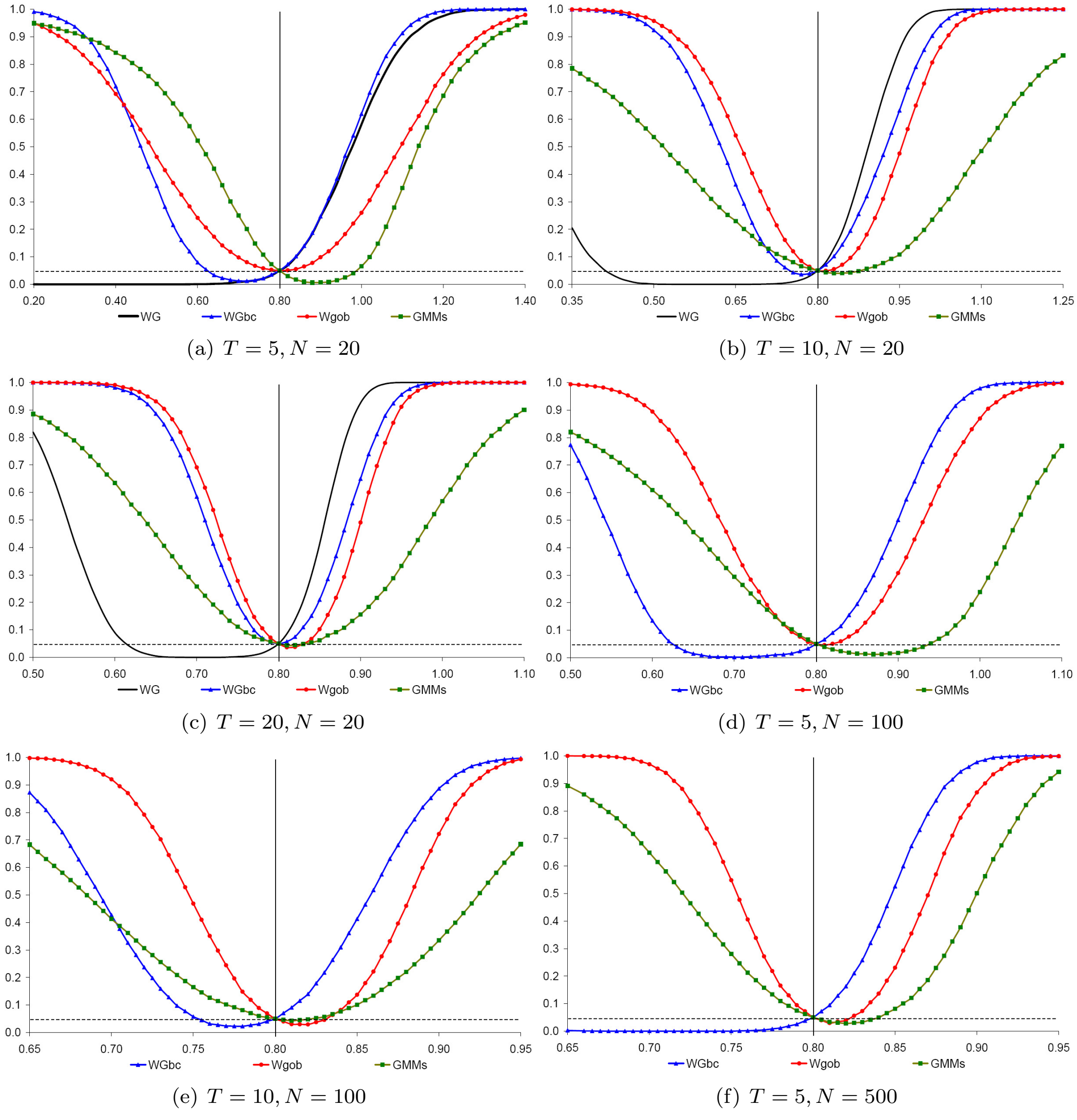
To conduct a power comparison, we compute the frequency of rejecting $H_0 : \theta = \theta_0$ using a 5% size-corrected critical value which is obtained from the size comparison conducted previously. Figure 2 contains power curves for the case $\theta = 0.8$, $\mu_\alpha = 5$ for a number of different sample sizes.[3] Because of the dramatic size distortions we do not report power curves for the WG estimator for $N = 100, 500$. We also do not report power curves for the GMMd estimator as it is outperformed by the GMMs estimator. The overall conclusion from these graphs is that the WGob estimator has superior power properties. First, the WG estimator has extremely low power for all of the considered sample sizes. Second, GMMs has better power properties but is still inferior to the WGob estimator in all cases, even for $N = 500$. Finally, WGbc is the only estimator with comparable, or even slightly better, power in the specific cases where T grows large. However, it has (extremely) low power for small values of T especially when N grows large.

[3] Results for the other cases are qualitatively comparable. They are available on request.

Table 2: Observed sizes (in percentages) of a two-sided t-statistic for testing the null hypothesis $\theta = \theta_0$ against the alternative $\theta \neq \theta_0$ using a 5% nominal size

T	5	10	20	5	10	5	5	10	20	5	10	5
N	20	20	20	100	100	500	20	20	20	100	100	500
Experiment 1: $\rho = 0.4,\ \sigma_s^2 = 2.0,\ \mu_\alpha = 1.0,\ \mu_\omega = 1.0,\ \gamma = 0.0$												
	$\theta = 0.4$						$\theta = 0.8$					
WG	44.0	28.1	16.2	98.4	87.2	100.0	97.6	94.5	81.4	100.0	100.0	100.0
WGbc	5.9	6.5	5.6	7.9	5.9	9.7	15.3	13.0	10.7	46.4	23.4	94.3
Wgob	8.4	8.6	8.5	6.3	6.4	8.9	9.4	8.6	8.8	6.4	6.4	6.2
GMMd	11.6	7.7	7.5	6.5	5.4	4.2	34.3	7.4	7.6	11.8	5.1	5.3
GMMs	12.7	8.4	8.5	6.7	5.7	5.4	8.6	9.3	8.2	6.6	5.4	6.9
Experiment 2: $\rho = 0.4,\ \sigma_s^2 = 2.0,\ \mu_\alpha = 1.0,\ \mu_\omega = 1.0,\ \gamma = 1.0$												
	$\theta = 0.4$						$\theta = 0.8$					
WG	44.0	28.1	16.2	98.4	87.2	100.0	97.6	94.5	81.4	100.0	100.0	100.0
WGbc	5.8	6.6	5.5	7.7	5.9	9.6	14.8	13.0	10.6	45.6	23.8	94.3
Wgob	8.3	8.7	9.4	6.7	8.6	11.4	9.4	8.7	9.0	6.4	6.5	6.4
GMMd	11.5	7.7	7.6	6.9	5.6	4.6	34.3	7.3	7.4	12.0	5.2	5.5
GMMs	19.0	9.3	8.6	8.7	6.0	5.7	8.2	9.2	8.3	6.8	5.5	7.0
Experiment 3: $\rho = 0.8,\ \sigma_s^2 = 2.0,\ \mu_\alpha = 1.0,\ \mu_\omega = 1.0,\ \gamma = 1.0$												
	$\theta = 0.4$						$\theta = 0.8$					
WG	63.1	39.3	22.4	99.9	96.6	100.0	98.2	95.8	83.7	100.0	100.0	100.0
WGbc	6.5	7.6	6.1	9.9	6.9	13.5	18.1	13.4	11.4	48.6	24.4	97.2
Wgob	8.0	9.0	9.6	6.3	9.5	9.8	8.9	8.1	8.5	6.0	6.4	6.0
GMMd	14.0	7.8	7.8	6.9	5.3	5.6	35.4	7.8	7.7	12.3	5.2	6.2
GMMs	10.5	9.4	8.6	9.9	6.5	7.0	8.2	8.9	7.6	6.3	5.6	5.7
Experiment 4: $\rho = 0.4,\ \sigma_s^2 = 2.0,\ \mu_\alpha = 5.0,\ \mu_\omega = 1.0,\ \gamma = 1.0$												
	$\theta = 0.4$						$\theta = 0.8$					
WG	44.0	28.1	16.2	98.4	87.2	100.0	97.6	94.5	81.4	100.0	100.0	100.0
WGbc	5.3	6.5	5.4	8.8	5.8	9.4	9.2	12.3	10.6	38.1	22.6	94.8
Wgob	8.2	10.4	12.1	8.8	18.0	22.3	9.5	9.3	10.2	6.7	8.9	8.0
GMMd	11.5	7.4	7.2	6.5	5.3	5.3	36.9	7.1	7.6	13.8	5.0	7.0
GMMs	65.9	12.5	9.3	24.7	6.4	7.9	21.6	10.6	8.8	14.2	6.0	7.9
Experiment 5: $\rho = 0.4,\ \sigma_s^2 = 2.0,\ \mu_\alpha = 1.0,\ \mu_\omega = 5.0,\ \gamma = 1.0$												
	$\theta = 0.4$						$\theta = 0.8$					
WG	44.0	28.1	16.2	98.4	87.2	100.0	97.6	94.5	81.4	100.0	100.0	100.0
WGbc	5.7	6.5	5.7	7.5	5.9	9.5	15.0	13.0	10.6	46.4	23.4	94.4
Wgob	8.3	8.2	7.7	6.1	5.8	8.0	9.2	8.7	8.3	6.1	5.9	6.4
GMMd	11.4	7.6	7.2	6.5	5.7	4.5	34.3	7.2	7.1	11.9	5.3	5.6
GMMs	13.0	8.0	8.6	7.8	6.1	6.1	7.3	9.2	8.0	6.3	5.0	6.8
Experiment 6: $\rho = 0.4,\ \sigma_s^2 = 8.0,\ \mu_\alpha = 1.0,\ \mu_\omega = 1.0,\ \gamma = 1.0$												
	$\theta = 0.4$						$\theta = 0.8$					
WG	19.3	12.8	8.9	69.6	45.1	100.0	70.7	56.8	37.9	100.0	99.8	100.0
WGbc	4.7	5.2	5.2	5.5	5.0	5.8	5.0	7.5	7.2	10.2	8.3	28.5
Wgob	8.5	8.3	8.3	6.4	7.0	8.6	8.4	8.5	8.4	6.1	5.4	6.8
GMMd	9.7	7.8	7.6	6.1	5.3	4.8	14.5	7.5	7.5	7.3	5.0	4.8
GMMs	12.5	8.1	8.6	7.1	5.7	5.7	7.4	7.2	7.8	5.3	4.9	4.8

Figure 2: Size-corrected power curves of a two-sided t-statistic for testing the null hypothesis $\theta = \theta_0$ against the alternative $\theta \neq \theta_0$ for the case $\theta = 0.8$, $\mu_\alpha = 5$



(a) $T = 5, N = 20$ (b) $T = 10, N = 20$

(c) $T = 20, N = 20$ (d) $T = 5, N = 100$

(e) $T = 10, N = 100$ (f) $T = 5, N = 500$

7 Conclusion

Dynamic panel data models are typically estimated using GMM. These instrumental variables estimators may exhibit serious small sample biases and/or relatively large standard deviations, especially in case of weak instruments. Bias-corrected WG estimators perform remarkably better in many cases, but the remaining bias may still be substantial when T is relatively small. Moreover, they are not always that straightforward to implement. In this paper, we retain a within-type of transformation but remove the individual effect from the lagged dependent variable by taking orthogonal deviations from its individual backward mean instead of from its individual sample mean. In the Hausman-Taylor approach, this transformed lagged dependent variable can be used as an instrument for the lagged dependent variable in the original model while for additional explanatory variables instruments can be selected depending on their alleged correlation with the individual effects and the idiosyncratic error term. This alternative estimator, referred to as WGob, is consistent for $T \to \infty$ but inconsistent for $N \to \infty$. However, the inconsistency is shown to be negligibly small. Moreover, a Monte Carlo simulation shows that this estimator is surprisingly accurate in comparison to established estimators. It considerably outperforms standard estimators in terms of bias, dispersion and inference in the cases where these estimators are know to fail, while not performing much worse in all other cases.

Acknowledgements

I thank Bart Cockx and conference and seminar participants in Bonn, Leuven, Muenster and Rotterdam for helpful suggestions and constructive comments on an earlier version of this paper. I acknowledge financial support from the Interuniversity Attraction Poles Program - Belgian Science Policy, contract no. P5/21.

References

Alvarez, J., and M. Arellano (2003): "The Time Series and Cross-Section Asymptotics of Dynamic Panel Data Estimators," *Econometrica*, 71, 1121–1159.

Anderson, T., and C. Hsiao (1981): "Estimation of Dynamic Models with Error Components," *Journal of the American Statistical Association*, 76, 598–606.

Anderson, T., and C. Hsiao (1982): "Formulation and estimation of dynamic models using panel data," *Journal of Econometrics*, 18, 47–82.

Arellano (2003): *Panel data econometrics*, Advanced texts in Econometrics. Oxford University Press.

Arellano, M., and S. Bond (1991): "Some Tests of Specification for Panel Data: Monte Carlo Evidence and an Application to Employment Equations," *Review of Economic Studies*, 58, 277–297.

Arellano, M., and O. Bover (1995): "Another look at the Instrumental Variable estimation of Error-Components Models," *Journal of Econometrics*, 68, 29–51.

Blundell, R., and S. Bond (1998): "Initial Conditions and Moment Restrictions in Dynamic Panel Data Models," *Journal of Econometrics*, 87, 115–143.

Bond, S., and F. Windmeijer (2005): "Reliable inference for GMM estimators? Finite sample properties of alternative test procedures in linear panel data models," *Econometric Reviews*, 24(1), 1–37.

Bun, M. (2003): "Bias correction in the dynamic panel data model with a nonscalar disturbance covariance matrix," *Econometric Reviews*, 22, 29–58.

Bun, M., and M. Carree (2005): "Bias-Corrected Estimation in Dynamic Panel Data Models," *Journal of Business and Economic Statistics*, 23(2), 200–210.

Bun, M., and J. Kiviet (2006): "The effects of dynamic feedbacks on LS and MM estimator accuracy in panel data models," *Journal of Econometrics*, 132(2), 409–444.

Bun, M. J. G., and F. Windmeijer (2010): "The Weak Instrument Problem of the System GMM Estimator in Dynamic Panel Data Models," *Econometrics Journal*, 13(1), 95–126.

Everaert, G., and L. Pozzi (2007): "Bootstrap-based bias correction for dynamic panels," *Journal of Economic Dynamics and Control*, 31(4), 1160–1184.

Hausman, J., and W. Taylor (1981): "Panel Data and Unobservable Individual Effects," *Econometrica*, 49, 1377–1398.

Judson, A., and A. Owen (1999): "Estimating Dynamic Panel Data Models: A Guide for Macroeconomists," *Economics Letters*, 65, 9–15.

Kiviet, J. (1995): "On Bias, Inconsistency, and Efficiency of Various Estimators in Dynamic Panel Data Models," *Journal of Econometrics*, 68, 53–78.

——— (2006): "Judging Contending Estimators by Simulation: Tournaments in Dynamic Panel Data Models," in *The Refinement of Econometric Estimation and Test Procedures*, ed. by G. Phillips, and E. Tzavalis. Cambridge University Press.

Nickell, S. (1981): "Biases in Dynamic Models with Fixed Effects," *Econometrica*, 49(6), 1417–1426.

Windmeijer, F. (2005): "A finite sample correction for the variance of linear efficient two-step GMM estimators," *Journal of Econometrics*, 126, 25–51.

Ziliak, J. (1997): "Efficient Estimation with Panel Data when Instruments are Predetermined: an Empirical Comparison of Moment-Condition Estimators," *Journal of Business and Economic Statistics*, 14(4), 419–431.

Appendices

Appendix A Proofs

Proof of Lemma 1. First note that by continuous substitution and using assumptions (A1)-(A3), we have from (1)

$$y_{i,t-1} = \frac{\alpha_i}{1-\theta} + \sum_{j=0}^{\infty} \theta^j \varepsilon_{i,t-j-1}, \tag{A-1}$$

$$\bar{y}^b_{i,t-1} = \frac{\alpha_i}{1-\theta} + \frac{1}{t}\sum_{s=1}^{t}\sum_{j=0}^{\infty} \theta^j \varepsilon_{i,s-j-1}. \tag{A-2}$$

Next, let $w_{i,t-1}$ be the homogeneous AR(1) process

$$w_{i,t-1} = y_{i,t-1} - \frac{\alpha_i}{1-\theta}, \tag{A-3}$$

with corresponding vectors $w_{-1} = \left(w'_{1,-1}, \ldots, w'_{N,-1}\right)'$ and $w_{i,-1} = \left(w_{i0}, \ldots, w_{i,T-1}\right)'$ such that $M^b_T w_i = \bar{w}^b_i$ and $Q^b_T w_i = \widetilde{w}^b_i = \widetilde{y}^b_i$.

Proof of (24): Letting ι_{1t} be a $T \times 1$ vector with the (s)th element being 1 for $s \leq t$ and 0 otherwise, such that $\overline{y}^b_{i,t-1} = \frac{1}{t}\iota'_{1t} y_{i,-1}$, we have

$$\begin{aligned}
\sigma^2_{\bar{y}^b,T} &= E\left(\frac{\bar{y}^{b\prime}_{-1}\bar{y}^b_{-1}}{NT}\right) = \frac{1}{T}E\left(\bar{y}^{b\prime}_{i,-1}\bar{y}^b_{i,-1}\right) = \frac{1}{T}\sum_{t=1}^{T}\frac{1}{t^2}\iota'_{1t}E\left(y_{i,-1}y'_{i,-1}\right)\iota_{1t}, \\
&= \frac{1}{T}\sum_{t=1}^{T}\frac{1}{t^2}\iota'_{1t}\left(\frac{\sigma^2_\alpha}{(1-\theta)^2}\iota_T\iota'_T + \frac{\sigma^2_\varepsilon}{1-\theta^2}\Sigma_T\right)\iota_{1t}, \\
&= \frac{\sigma^2_\alpha}{(1-\theta)^2} + \frac{\sigma^2_\varepsilon}{(1-\theta)^2}\sum_{t=1}^{T}\frac{1}{t}\left(1 - \frac{2\theta}{t}\frac{1-\theta^t}{1-\theta^2}\right) = \frac{\sigma^2_\alpha}{(1-\theta)^2} + O\left(\frac{\log T}{T}\right),
\end{aligned} \tag{A-4}$$

where Σ_T is a $T \times T$ symmetric matrix with the (t,s)th element being $\theta^{|t-s|}$. The order of approximation is obtained from using

$$\lim_{T\to\infty}\sum_{t=1}^{T}\frac{1}{t} = \gamma + \ln(T), \qquad \lim_{T\to\infty}\sum_{t=1}^{T}\frac{\theta^t}{t} = -\ln(1-\theta), \qquad \lim_{T\to\infty}\sum_{t=1}^{T}\frac{1}{t^2} = \frac{\pi^2}{6}, \tag{A-5}$$

with γ being the Euler-Mascheroni constant and from noting that the dilogarithm

$$\lim_{T\to\infty}\sum_{t=1}^{T}\theta^t/t^2 \equiv \mathrm{Li}_2(\theta), \tag{A-6}$$

cannot be evaluated in closed form for all values of θ but is convergent for the relevant range $-1 < \theta < 1$.

Proof of (25): Letting ι_t be a $T \times 1$ vector with the (t)th element being 1 and 0 otherwise, such that $w_{i,t-1} = \iota_t' w_{i,-1}$, we have from the definition of $w_{i,t}$

$$
\begin{aligned}
\sigma_{y\overline{y}^b,T} &= E\left(\frac{y_{-1}'\overline{y}_{-1}^b}{NT}\right) = \frac{1}{T}E\left(y_{i,-1}'\overline{y}_{i,-1}^b\right) = \frac{1}{T}\sum_{t=1}^{T} E\left(y_{i,t-1}\overline{y}_{i,t-1}^b\right),\\
&= \frac{1}{T}\sum_{t=1}^{T} E\left(\left(\frac{\alpha_i}{1-\theta} + w_{i,t-1}\right)\left(\frac{\alpha_i}{1-\theta} + \overline{w}_{i,t-1}^b\right)\right),\\
&= \frac{\sigma_\alpha^2}{(1-\theta)^2} + \frac{1}{T}\sum_{t=1}^{T} E\left(w_{i,t-1}\overline{w}_{i,t-1}^b\right) = \frac{\sigma_\alpha^2}{(1-\theta)^2} + \frac{1}{T}\sum_{t=1}^{T}\frac{1}{t}\iota_t' E\left(w_{i,-1}w_{i,-1}'\right)\iota_{1t},\\
&= \frac{\sigma_\alpha^2}{(1-\theta)^2} + \frac{\sigma_\varepsilon^2}{1-\theta^2}\frac{1}{T}\sum_{t=1}^{T}\frac{1}{t}\iota_t'\Sigma_T\iota_{1t} = \frac{\sigma_\alpha^2}{(1-\theta)^2} + \frac{\sigma_\varepsilon^2}{1-\theta^2}\frac{1}{T}\sum_{t=1}^{T}\frac{1}{t}\frac{1-\theta^t}{1-\theta},\\
&= \frac{\sigma_\alpha^2}{(1-\theta)^2} + O\left(\frac{\log T}{T}\right). \qquad \text{(A-7)}
\end{aligned}
$$

Proof of (26): From the definition of $w_{i,t}$ we have

$$
\begin{aligned}
\sigma_{\overline{y}^b\widetilde{y}^b,T} &= E\left(\frac{\bar{y}_{-1}^{b'}\widetilde{y}_{-1}^b}{NT}\right) = \frac{1}{T}E\left(\bar{y}_{i,-1}^{b'}\widetilde{y}_{i,-1}^b\right) = \frac{1}{T}E\left(\bar{y}_{i,-1}^{b'}\widetilde{w}_{i,-1}^b\right),\\
&= \frac{1}{T}E\left(\left(\frac{\alpha_i}{1-\theta} + \bar{w}_{i,-1}^{b'}\right)\widetilde{w}_{i,-1}^b\right) = \frac{1}{T}E\left(\bar{w}_{i,-1}^{b'}\widetilde{w}_{i,-1}^b\right),\\
&= \frac{1}{T}\left(E\left(\bar{w}_{i,-1}^{b'}w_{i,-1}\right) - E\left(\bar{w}_{i,-1}^{b'}\bar{w}_{i,-1}^b\right)\right),\\
&= \frac{\sigma_\varepsilon^2}{1-\theta^2}\frac{1}{T}\sum_{t=1}^{T}\frac{1}{t}\frac{1-\theta^t}{1-\theta} - \frac{\sigma_\varepsilon^2}{1-\theta^2}\frac{1}{T}\sum_{t=1}^{T}\frac{1}{t}\left(\frac{1+\theta}{1-\theta} - \frac{2\theta}{t}\frac{1-\theta^t}{(1-\theta)^2}\right),\\
&= -\frac{\theta}{1+\theta}\frac{\sigma_\varepsilon^2}{(1-\theta)^2}\frac{1}{T}\sum_{t=1}^{T}\frac{1}{t}\left(1+\theta^{t-1} - \frac{2}{t}\frac{1-\theta^t}{1-\theta}\right) = O\left(\frac{\log T}{T}\right). \qquad \text{(A-8)}
\end{aligned}
$$

Proof of Theorem 1. Using the results in Lemma 1, the inconsistency of the WGob estimator in (23) can be written as

$$
\begin{aligned}
\operatorname*{plim}_{N\to\infty}\left(\widehat{\theta}_{\perp}^{WG}-\theta\right) &= \frac{1-\delta_T}{\frac{\sigma_\alpha^2}{(1-\theta)^2}+\frac{\sigma_\varepsilon^2}{1-\theta^2}-\delta_T\left(\frac{\sigma_\alpha^2}{(1-\theta)^2}+\frac{\sigma_\varepsilon^2}{1-\theta^2}\frac{1}{T}\sum_{t=1}^{T}\frac{1}{t}\frac{1-\theta^t}{1-\theta}\right)}\frac{\sigma_\alpha^2}{1-\theta}, \\
&= \frac{1-\delta_T}{(1-\delta_T)\frac{\sigma_\alpha^2}{(1-\theta)^2}+\left(1-\delta_T\frac{1}{T}\sum_{t=1}^{T}\frac{1}{t}\frac{1-\theta^t}{1-\theta}\right)\frac{\sigma_\varepsilon^2}{1-\theta^2}}\frac{\sigma_\alpha^2}{1-\theta}, \\
&= \frac{1-\theta}{1+\frac{1-\theta}{1+\theta}\left(1+\left(1-\frac{1}{T}\sum_{t=1}^{T}\frac{1}{t}\frac{1-\theta^t}{1-\theta}\right)\frac{\delta_T}{1-\delta_T}\right)\frac{\sigma_\varepsilon^2}{\sigma_\alpha^2}}, \\
&= \frac{(1-\theta)}{1+\frac{1-\theta}{1+\theta}\left(1+\left(1-\frac{1}{T}\sum_{t=1}^{T}\frac{1}{t}\frac{1-\theta^t}{1-\theta}\right)\frac{1+\frac{1}{T}\sum_{t=1}^{T}\frac{1}{t}\frac{1-\theta^t}{1+\theta}\frac{\sigma_\varepsilon^2}{\sigma_\alpha^2}}{\frac{1}{T}\sum_{t=1}^{T}\frac{1}{t}\left(1-\frac{2\theta}{t}\frac{1-\theta^t}{1-\theta^2}-\frac{1-\theta^t}{1+\theta}\right)\frac{\sigma_\varepsilon^2}{\sigma_\alpha^2}}\right)\frac{\sigma_\varepsilon^2}{\sigma_\alpha^2}}, \\
&= \frac{(1-\theta)}{1+\frac{1-\theta}{1+\theta}\frac{\sigma_\varepsilon^2}{\sigma_\alpha^2}+\frac{(1-\theta)\left(1-\frac{1}{T}\sum_{t=1}^{T}\frac{1}{t}\frac{1-\theta^t}{1-\theta}\right)\left(1+\frac{1}{T}\sum_{t=1}^{T}\frac{1}{t}\frac{1-\theta^t}{1+\theta}\frac{\sigma_\varepsilon^2}{\sigma_\alpha^2}\right)}{\theta\frac{1}{T}\sum_{t=1}^{T}\frac{1}{t}\left(1+\theta^{t-1}-\frac{2}{t}\frac{1-\theta^t}{1-\theta}\right)}}, \\
&= \frac{(1-\theta)}{1+\frac{1-\theta}{1+\theta}\frac{\sigma_\varepsilon^2}{\sigma_\alpha^2}+\frac{((1-\theta)-B_T)\left(1+\frac{1}{T}\sum_{t=1}^{T}\frac{1}{t}\frac{1-\theta^t}{1+\theta}\frac{\sigma_\varepsilon^2}{\sigma_\alpha^2}\right)}{\theta A_T}}, \\
&= \frac{\theta(1-\theta)A_T}{\theta A_T\left(1+\frac{1-\theta}{1+\theta}\frac{\sigma_\varepsilon^2}{\sigma_\alpha^2}\right)+((1-\theta)-B_T)\left(1+\frac{1}{T}\sum_{t=1}^{T}\frac{1}{t}\frac{1-\theta^t}{1+\theta}\frac{\sigma_\varepsilon^2}{\sigma_\alpha^2}\right)}, \\
&= \frac{\theta(1-\theta)A_T}{(1-\theta)+\theta A_T-B_T+\frac{\sigma_\varepsilon^2}{\sigma_\alpha^2}C_T},
\end{aligned}
$$

where

$$
\begin{aligned}
A_T &= \frac{1}{T}\sum_{t=1}^{T}\frac{1}{t}\left(1+\theta^{t-1}-\frac{2}{t}\frac{1-\theta^t}{1-\theta}\right), \\
B_T &= \frac{1}{T}\sum_{t=1}^{T}\frac{1}{t}\left(1-\theta^t\right), \\
C_T &= (1-\theta)\frac{1}{T}\sum_{t=1}^{T}\frac{1}{t}\left(1-\frac{2}{t}\frac{1-\theta^t}{1-\theta^2}\right)-\frac{B_T^2}{1+\theta}.
\end{aligned}
$$

The approximation in (30) follows from A_T, B_T and C_T being $O\left(\frac{\log T}{T}\right)$.