UNIVERSITEIT GENT

FACULTEIT ECONOMIE
EN BEDRIJFSKUNDE

TWEEKERKENSTRAAT 2
B-9000 GENT
Tel. : 32 - (0)9 – 264.34.61
Fax. : 32 - (0)9 – 264.35.92

WORKING PAPER

Using Backward Means to Eliminate Individual Effects from Dynamic Panels

Gerdie Everaert

January 2009

2009/553

D/2009/7012/05

Using Backward Means to Eliminate Individual Effects from Dynamic Panels *

Gerdie Everaert†

January 23, 2009

Abstract

The within-groups estimator is inconsistent in dynamic panels with fixed T since the sample mean used to eliminate the individual effects from the lagged dependent variable is correlated with the error term. This paper suggests to eliminate individual effects from an AR(1) panel using backward means as an alternative to sample means. Using orthogonal deviations of the lagged dependent variable from its backward mean yields an estimator that is still inconsistent for fixed T but the inconsistency is shown to be negligibly small. A Monte Carlo simulation shows that this alternative estimator has superior small sample properties compared to conventional fixed effects, bias-corrected fixed effects and GMM estimators. Interestingly, it is also consistent for fixed T in the specific cases where (i) $T = 2$, (ii) the AR parameter is 0 or 1, (iii) the variance of the individual effects is zero.

JEL Classification: C15, C32

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

*I thank Bart Cockx 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.

†SHERPPA, Ghent University, Gerdie.Everaert@UGent.be, http://www.sherppa.be

1 Introduction

We know from Nickell (1981) that in dynamic panels with individual effects the fixed effects or within-groups (WG) estimator is inconsistent when the cross-sectional dimension N tends to infinity but the time dimension T is fixed. Given this inconsistency, the literature has focused mainly on 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 standard GMM estimators (i) have a (much) larger standard error compared to least squares (LS) estimators (see e.g. Arellano and Bond, 1991; Kiviet, 1995) and (ii) may suffer from substantial finite sample bias due to a weak instruments problem (see Ziliak, 1997; Bun and Kiviet, 2006). 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 be substantial when T is relatively small.

In this paper we follow a different route. We stick to LS estimation of the model but remove the individual effects from an AR(1) panel in a slightly different way than by taking deviations from sample means. Inspection of the WG estimator for $N \to \infty$ shows that its inconsistency stems from the asymptotic correlation between the within-transformed, i.e. in deviation from its individual sample mean, lagged dependent variable and the idiosyncratic error term at time t. This correlation is due to the fact that the sample mean of the lagged dependent variable includes observations for time $t, \ldots, T$ which are all affected by the idiosyncratic error term at time t. This suggests that obtaining a consistent LS estimator for $N \to \infty$ requires the variable that eliminates the individual effects used at time t to be orthogonal to the innovations at time $t, \ldots, T$. Therefore, this paper suggests to transform the lagged dependent variable into orthogonal deviations from its backward mean in stead of from its sample mean. This is equivalent to adding the backward mean of the lagged dependent variable as a regressor in the model, which then serves as a proxy for the individual effects. This alternative estimator, referred to as WGob, is shown to be consistent for $T \to \infty$ but converges at a slower rate and is inconsistent for $N \to \infty$ and T fixed. Fortunately, this inconsistency is shown to be negligibly small. Interestingly, the WGob estimator is consistent for fixed T in the specific cases where (i) $T = 2$; (ii) the AR(1) coefficient is either 0 or 1 and (iii) the ratio of the variance of individual effects over the variance of the idiosyncratic error is zero. Note that for small values of T and an AR(1) parameter close to 1 standard estimators are known to fail.

A Monte Carlo simulation is used to examine the finite sample properties of the WGob estimator compared to first-differenced and system GMM estimators and the bias-corrected WG estimator suggested by Kiviet (1995). The most important finding is that, despite being inconsistent for $N \to \infty$ and converging at a slower rate when $T \to \infty$, the WGob estimator seems more attractive than the standard estimators as it is found to be more robust with respect to alternative parameter values. As such it considerably outperforms the standard estimators in terms of bias and dispersion in the cases where these estimators are know to fail, while not performing much worse in all other cases.

The layout of the paper is as follows. Section 2 presents the model and the assumptions. Section 3 motivates using orthogonal deviations from backward means from inspection of the Nickell bias. Section 4 analyses the asymptotic properties of the WGob estimator. Section 5 presents the results of a Monte Carlo simulation comparing the finite sample performance of the suggested WGob estimator to a number of standard dynamic panel data estimators. Section 6 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 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 unobserved disturbance term ε_{it}. For notational convenience we assume y_{i0} is observed. We further assume:

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 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 in model (3) are cancelled out by premultiplying 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 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

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

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

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

where $c = diag\left[1, \sqrt{2}, \frac{\sqrt{t}}{\sqrt{t-1}}, \ldots, \frac{\sqrt{T}}{\sqrt{T-1}}\right]$, $\bar{y}_i^b = \left[\bar{y}_{i1}^b, \ldots, \bar{y}_{iT}^b\right]'$ 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 $\bar{y}_{it,-1}^b = t^{-1}\sum_{s=0}^{t-1} y_{is}$.

Note that opposed to Q, Q^b is not a symmetric and idempotent matrix. As the rows of Q_T^b add up to zero, i.e. $Q_T^b \iota_T = 0$, the individual effects in model (3) are cancelled out by premultiplying by Q^b, obtaining

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

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 $\widetilde{\varepsilon}^b = \left(\widetilde{\varepsilon}_1^{b'}, \ldots, \widetilde{\varepsilon}_N^{b'}\right)'$, $\widetilde{\varepsilon}_i^b = \left(\widetilde{\varepsilon}_{i1}^b, \ldots, \widetilde{\varepsilon}_{iT}^b\right)'$ with $\widetilde{\varepsilon}_{it}^b = \varepsilon_{it} - \bar{\varepsilon}_{it}^b$ and $\bar{\varepsilon}_{it}^b = t^{-1}\sum_{s=1}^t \varepsilon_{is}$. The scale factor c is introduced to ensure that $Q^b Q^{b'} = I$ such that the transformation preserves the orthogonality of the error terms, i.e. if $\text{var}(\varepsilon_i) = \sigma^2 I_T$ then $\widetilde{\varepsilon}_i^b$ also has $\text{var}\left(\widetilde{\varepsilon}_i^b\right) = \sigma^2 I_T$. The least squares estimate of θ in equation (11)

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

indeed equals the WG estimator in (7) as it can easily be verified that $Q^{b'}Q^b = Q$.

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 $QD = 0$

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

shows that this inconsistency stems from the fact that for fixed T the term $y_{-1}'Q\varepsilon$ 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 correlated with the error term ε_{it}. Obtaining a consistent LS estimator for $N \to \infty$ requires the variable that eliminates the individual effects used at time t to be orthogonal to the innovations $\varepsilon_{it}, \ldots, \varepsilon_{iT}$. This suggests using backward means in stead of sample means. However the representation of the WG estimator in (12) shows that this yields exactly the same estimator. Inserting (11) in (12)

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

shows that the inconsistency of the WG estimator can also be seen to stem from the correlation between y_{-1} and ε in deviation from their backward means. Interestingly, $\widetilde{y}_{it,-1}^b = y_{i,t-1} - \bar{y}_{i,-1}^b$ is not correlated with the error term ε_{it} but the inconsistency of $\widehat{\theta}^{WG}$ in (14) is due to the fact that Q^b is not symmetric and idempotent which implies that $y_{-1}'Q^{b'}Q^b\varepsilon \neq y_{-1}'Q^{b'}\varepsilon$. Therefore, the next section analyses the properties of an estimator that uses an alternative, symmetric and idempotent, backward mean operator.

For comparison with the estimator presented below, the expression for the inconsistency of the WG estimator for $N \to \infty$ is given by (see Nickell, 1981; Anderson and Hsiao, 1981)

$$\operatorname*{plim}_{N\to\infty}\left(\widehat{\theta}^{WG}-\theta\right) = -\frac{(1+\theta)\left(1-\frac{1}{T}\frac{1-\theta^T}{1-\theta}\right)}{T-1-\frac{2\theta}{1-\theta}\left(1-\frac{1}{T}\frac{1-\theta^T}{1-\theta}\right)}. \tag{15}$$

For small values of T the inconsistency is given by

$$\operatorname*{plim}_{N\to\infty}\left(\widehat{\theta}^{WG}-\theta\right) = -\frac{1+\theta}{2} \qquad \text{for} \quad T=2, \tag{16}$$

$$= -\frac{(1+\theta)(2+\theta)}{2(3+\theta)} \qquad \text{for} \quad T=3, \tag{17}$$

while for reasonably large values of T, (15) can be approximated by

$$\operatorname*{plim}_{N\to\infty}\left(\widehat{\theta}^{WG}-\theta\right) \simeq -\frac{1+\theta}{T}. \tag{18}$$

4 Orthogonalising regressors to backward means

The backward mean representation in (14) shows that the inconsistency of the WG estimator for $N \to \infty$ stems from the fact that Q^b is not a symmetric and idempotent matrix. Therefore, instead of taking deviations from backward means define the backward orthogonal operator $Q^b_\perp$

$$Q^b_\perp = I_{NT} - \bar{y}^b_{-1}\left(\bar{y}^{b\prime}_{-1}\bar{y}^b_{-1}\right)^{-1}\bar{y}^{b\prime}_{-1}, \tag{19}$$

where $\bar{y}^b_{-1} = \left(\bar{y}^{b\prime}_{1,-1},\dots,\bar{y}^{b\prime}_{N,-1}\right)'$ 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

$$\widehat{y}^b = \theta\widehat{y}^b_{-1} + \widehat{\alpha}^b + \widehat{\varepsilon}^b, \tag{20}$$

where $\widehat{y}^b$, $\widehat{y}^b_{-1}$, $\widehat{\alpha}^b$ and $\widehat{\varepsilon}^b$ are the residuals of the auxiliary regressions of y, y_{-1}, α and ε on $\bar{y}^b_{-1}$. The LS estimator for θ in (20), we shall refer to this as WGob, is given by

$$\widehat{\theta}^{WG}_\perp = \left(y'_{-1}Q^b_\perp y_{-1}\right)^{-1} y'_{-1}Q^b_\perp y = \left(\widehat{y}^{b\prime}_{-1}y_{-1}\right)^{-1}\widehat{y}^{b\prime}_{-1}y. \tag{21}$$

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

Remark 1. Using the Frisch-Waugh-Lovell theorem, $\widehat{\theta}^{WG}_\perp$ 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.

Inserting (3) in (21) yields

$$\widehat{\theta}_{\perp}^{WG} = \theta + \left(\widehat{y}_{-1}^{b\prime} y_{-1}\right)^{-1} \widehat{y}_{-1}^{b\prime} \left(D\alpha + \varepsilon\right) = \theta + \left(y_{-1}^{\prime} Q_{\perp}^{b} y_{-1}\right)^{-1} y_{-1}^{\prime} Q_{\perp}^{b} \left(D\alpha + \varepsilon\right), \tag{22}$$

In contrast to Q^b, the rows of $Q_{\perp}^b$ do not sum to zero by construction such that $Q_{\perp}^b D$ is not necessarily 0. This implies that by premultiplying the data by $Q_{\perp}^b$ the individual effects in α are not cancelled out exactly such that the transformed explanatory variable $\widehat{y}_{-1}^{b}$ in the numerator of (22) is potentially correlated with the error term $D\alpha + \varepsilon$.

The results collected in the following two Lemma's are useful to establish the asymptotic properties of the WGob estimator. All proofs are in the appendix.

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

$$\sigma_{y_{-1}}^2 = \frac{1}{NT} E\left[y_{-1}^{\prime} y_{-1}\right] = \frac{\sigma_\alpha^2}{(1-\theta)^2} + \frac{\sigma_\varepsilon^2}{1-\theta^2}, \tag{23}$$

$$\sigma_{\bar{y}_{-1}^b}^2 = \frac{1}{NT} E\left[\bar{y}_{-1}^{b\prime} \bar{y}_{-1}^{b}\right] = \frac{\sigma_\alpha^2}{(1-\theta)^2} + \frac{\sigma_\varepsilon^2}{(1-\theta)^2} \frac{1}{T} \sum_{t=1}^{T} \frac{1}{t} \left(1 - \frac{2\theta}{t} \frac{1-\theta^t}{1-\theta^2}\right), \tag{24}$$

$$\sigma_{y_{-1}\bar{y}_{-1}^b} = \frac{1}{NT} E\left[y_{-1}^{\prime} \bar{y}_{-1}^{b}\right] = \frac{\sigma_\alpha^2}{(1-\theta)^2} + \frac{\sigma_\varepsilon^2}{1-\theta^2} \frac{1}{T} \sum_{t=1}^{T} \frac{1}{t} \frac{1-\theta^t}{1-\theta}, \tag{25}$$

where $\sigma_{y_{-1}}^2$ is the population variance of y_{-1}, $\sigma_{\bar{y}_{-1}^b}^2$ is the population variance of $\bar{y}_{-1}^b$ and $\sigma_{y_{-1}\bar{y}_{-1}^b}$ is the population covariance between y_{-1} and $\bar{y}_{-1}^b$.

Lemma 2. *For $T \to \infty$ we have*

$$\underset{T\to\infty}{plim} \left(\bar{y}_{-1}^{b\prime} \bar{y}_{-1}^{b}\right)^{-1} \bar{y}_{-1}^{b\prime} y_{-1} = 1, \tag{26}$$

such that

$$\underset{T\to\infty}{plim}\, \widehat{y}_{-1}^{b} = \underset{T\to\infty}{plim}\, Q_{\perp}^{b} y_{-1} = y_{-1} - \bar{y}_{-1}^{b} = \widetilde{y}_{-1}^{b} = Q^b y_{-1}. \tag{27}$$

Moreover

$$\frac{1}{NT} E\left[\widetilde{y}_{-1}^{b\prime} \left(D\alpha + \varepsilon_{it}\right)\right] = 0, \tag{28}$$

$$\frac{1}{NT} E\left[\widetilde{y}_{-1}^{b\prime} y_{-1}\right] = \frac{\sigma_\varepsilon^2}{1-\theta^2} \left(1 - \frac{1}{T} \sum_{t=1}^{T} \frac{1}{t} \frac{1-\theta^t}{1-\theta}\right). \tag{29}$$

For $N \to \infty$ we have

$$\underset{N\to\infty}{plim} \frac{1}{NT}\widehat{y}^{b'}_{-1}(\alpha D + \varepsilon) = \left(1 - \delta_{y_{-1}\bar{y}^b_{-1}}\right)\frac{\sigma^2_\alpha}{1-\theta}, \tag{30}$$

$$\underset{N\to\infty}{plim} \frac{1}{NT}\widehat{y}^{b'}_{-1}\widehat{y}^b_{-1} = \left(1 - \rho^2_{y_{-1}\bar{y}^b_{-1}}\right)\sigma^2_{y_{-1}}, \tag{31}$$

where $\rho_{y_{-1}\bar{y}^b_{-1}} = \sigma_{y_{-1}\bar{y}^b_{-1}} \Big/ \left(\sigma_{t,\bar{y}^b_{-1}}\sigma_{y_{-1}}\right)$ is the population correlation between $y_{i,t-1}$ and $\bar{y}^b_{it,-1}$ and $\delta_{y_{-1}\bar{y}^b_{-1}} = \sigma_{y_{-1}\bar{y}^b_{-1}} \Big/ \sigma^2_{\bar{y}^b_{-1}}$ is the population OLS regression coefficient of y_{-1} on $\bar{y}^b_{-1}$.

The first part of this Lemma shows that as $T \to \infty$, $\widehat{y}^b_{-1}$ converges to $\widetilde{y}^b_{-1}$ which is uncorrelated with $D\alpha + \varepsilon$. The second part provides the variance of $\widehat{y}^b_{-1}$ and its covariance with $D\alpha + \varepsilon$ when $N \to \infty$. This allows us to derive the asymptotic properties of $\widehat{\theta}^{WG}_\perp$.

Theorem 1. *$\widehat{\theta}^{WG}_\perp$ is consistent as $T \to \infty$ (regardless of whether N is fixed or tends to infinity) but inconsistent as $N \to \infty$ (and T fixed) with the asymptotic bias term given by*

$$\underset{N\to\infty}{plim}\left(\widehat{\theta}^{WG}_\perp - \theta\right) = \frac{1 - \delta_{y_{-1}\bar{y}^b_{-1}}}{\left(1 - \rho^2_{y_{-1}\bar{y}^b_{-1}}\right)\sigma^2_{y_{-1}}}\frac{\sigma^2_\alpha}{1-\theta}, \tag{32}$$

$$= \frac{\theta(1-\theta)A_T}{(1-\theta) + B_T + C_T\frac{\sigma^2_\varepsilon}{\sigma^2_\alpha}}, \tag{33}$$

where

$$A_T = \frac{1}{T}\sum_{t=1}^{T}\frac{1}{t}\left(1 + \theta^{t-1} - \frac{2}{t}\frac{1-\theta^t}{1-\theta}\right), \tag{34}$$

$$B_T = \frac{1}{T}\sum_{t=1}^{T}\frac{1}{t}\left(2\theta^t - (1-\theta) - \frac{2\theta}{t}\frac{1-\theta^t}{1-\theta}\right), \tag{35}$$

$$C_T = \frac{1}{T}\sum_{t=1}^{T}\frac{1}{t}\left((1-\theta) - \frac{2\theta}{t}\frac{1-\theta^t}{1+\theta} - \frac{1}{t}\frac{1}{T}\frac{\left(1-\theta^t\right)^2}{1+\theta}\right). \tag{36}$$

This theorem shows that the asymptotic properties of the WGob estimator are qualitatively the same as those of the standard WG estimator. The following corollary provides a quantitative analysis of the asymptotic bias of $\widehat{\theta}^{WG}_\perp$ for $N \to \infty$.

Corollary 1.

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

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

$$\underset{N,\frac{\sigma^2_\alpha}{\sigma^2_\varepsilon}\to\infty}{plim}\left(\widehat{\theta}^{WG}_\perp - \theta\right) = \frac{\theta(1-\theta)\sum_{t=1}^{T}\frac{1}{t}\left(1 + \theta^{t-1} - \frac{2}{t}\frac{1-\theta^t}{1-\theta}\right)}{T(1-\theta) + \sum_{t=1}^{T}\frac{1}{t}\left(2\theta^t - (1-\theta) - \frac{2\theta}{t}\frac{1-\theta^t}{1-\theta}\right)}, \tag{37}$$

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

$$\underset{N\to\infty}{plim}\left(\widehat{\theta}_{\perp}^{WG}-\theta\right)=0 \qquad for \quad T=2, \tag{38}$$

$$\underset{N,\frac{\sigma_{\alpha}^{2}}{\sigma_{\varepsilon}^{2}}\to\infty}{plim}\left(\widehat{\theta}_{\perp}^{WG}-\theta\right)=\frac{\theta\left(1-\theta\right)}{4\left(3-3/8+\theta\right)} \qquad for \quad T=3, \tag{39}$$

while for reasonably large vales of T, *(33) and* $\theta<1$ *can be approximated by*

$$\underset{N\to\infty}{plim}\left(\widehat{\theta}_{\perp}^{WG}-\theta\right)\simeq\theta\left(1-\theta\right)\frac{\ln\left(T\right)}{T}. \tag{40}$$

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

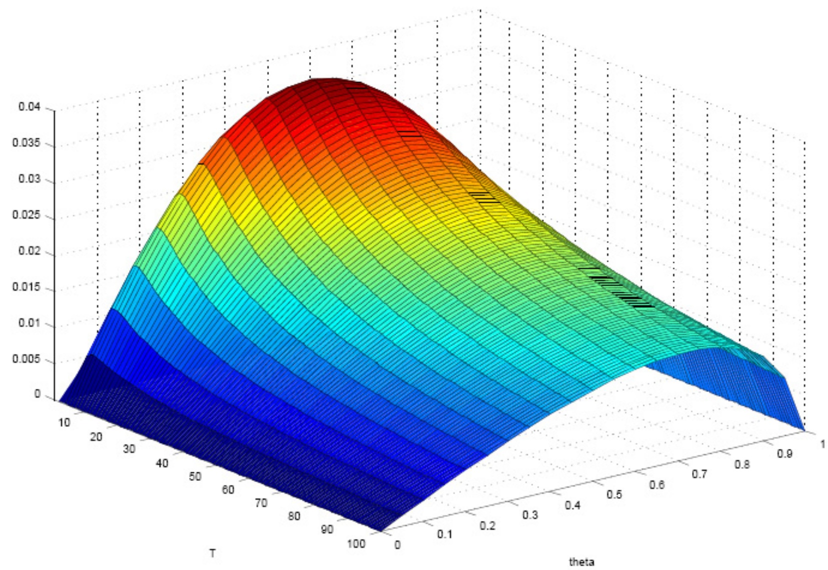
Interestingly, comparing (16)-(18) to (38)-(40) shows that, over the relevant range $0\leq\theta\leq1$, the inconsistency of $\widehat{\theta}_{\perp}^{WG}$ for fixed T is much smaller than that of the WG estimator. Moreover, $\widehat{\theta}_{\perp}^{WG}$ is consistent for $N\to\infty$ and fixed T when T is extremely small, i.e. $T=2$, when θ is either 0 or 1 and when $\sigma_{\alpha}^{2}=0$. 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, calculated from (37), 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. This suggests that this alternative estimator may be of great practical relevance. 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 Monte Carlo study of finite sample properties

5.1 Design

In this section we analyse the small sample performance of the WGob estimator presented above using a Monte Carlo simulation. To generate data from (1) we make a number of additional assumptions. First, we make the distributional assumptions $\varepsilon_{it}\sim N\left(0,\sigma_{\varepsilon}^{2}\right)$ and $\alpha_{i}\sim N\left(0,\sigma_{\alpha}^{2}\right)$. Second, y_{i0} is drawn from its stationary distribution $N\left(\frac{\alpha_i}{1-\theta},\frac{\sigma_{\varepsilon}^{2}}{1-\theta^{2}}\right)$. Third, we impose the normalisation restriction $\sigma_{\varepsilon}^{2}=1$ and calculate $\sigma_{\alpha}=\mu\sigma_{\varepsilon}\left(1-\theta\right)$ where the value of μ controls the relative impact on y_{it} of the disturbance ε_{it} versus the individual effect α_{i}. The performance

Figure 1: Inconsistency of $\widehat{\theta}_{\perp}^{WG}$ for $N \to \infty$ and $\frac{\sigma_\alpha^2}{\sigma_\varepsilon^2} \to \infty$



of the WGob estimator under different parameter combinations is compared to 4 alternative dynamic panel data estimators: (i) WG, the standard within groups estimator, (ii) GMMd, the first difference GMM estimator proposed by Arellano and Bond (1991), (iii) GMMs, the system GMM estimator proposed by Arellano and Bover (1995) and Blundell and Bond (1998), (iv) WGbc, the bias-corrected WG estimator proposed by Kiviet (1995). For the 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). 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 former is applicable in the proposed Monte Carlo design. To implement the WGbc estimator we use the GMMs estimator as an initial large-N consistent estimator. As due to a weak instruments problem the GMM estimators do not necessarily have first or second finite sample moments, we use the median bias (MB), the median absolute deviation (MAD) and the median absolute error (MAE) as measures to compare the different estimators.[1] We consider the following experiments: $\theta \in \{0.4, 0.8\}$, $\mu \in \{1, 5\}$, $(T, N) \in \{(2,20), (5,20), (10,20), (20,20), (40,20), (2,100), (5,100), (10,100), (2,500), (5,500)\}$. For each experiment, we performed 1000 Monte Carlo replications.

[1]The MAD is defined as the median of the absolute value of the deviation of an estimator from its median estimate over the Monte Carlo replications, while the MAE is the median of the absolute value of the deviation of an estimator from its population value.

5.2 Results

The results of the Monte Carlo simulation are presented in Tables 1-4. As known from the existing literature (i) the WG estimator is severely biased, especially when θ approaches 1, with the bias only disappearing as T grows larger and not as N increases (see Nickell, 1981), but has a dispersion smaller than achieved by any of the consistent estimators; (ii) 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 becomes large (see Blundell and Bond, 1998); (iii) 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 is large (see Kiviet, 2006); (iv) the WGbc estimator even 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 (severely) biased when T is small.

Turning to the WGob estimator, despite being inconsistent for $N \to \infty$ and converging at a slower rate when $T \to \infty$ it performs remarkably well in terms of median point estimates, i.e. in none of the considered cases it exhibits a considerable bias. As a result the WGob estimator clearly outperforms the other estimators in terms of MB in the cases where these estimators are (severely) biased while not being dramatically worse in the cases where these estimators are consistent. The dispersion of the WGob is larger than that of the WG and WGbc estimators but smaller than that of the GMM estimators in most cases. This implies that in terms of MAE, 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 $N = 500$.

On the whole, the WGob estimator seems more attractive than the GMM estimators and the bias-corrected WG estimator as its performance is found to be robust over the different experiments. As such it considerably outperforms the other estimators in terms of MB and MAE in a number of cases while being more or less comparable in all other cases.

Table 1: Monte Carlo comparison of alternative dynamic panel data estimators for θ when $\theta = 0.4, \mu = 1$

	T	N	MB	MAD	MAE	T	N	MB	MAD	MAE
	2	20				2	100			
WG			-0.696	0.150	0.696			-0.703	0.067	0.703
WGbc			-0.233	0.183	0.260			-0.200	0.095	0.200
GMMd			-0.065	0.442	0.472			0.005	0.206	0.206
GMMs			-0.001	0.270	0.271			-0.010	0.130	0.130
WGob			-0.026	0.294	0.294			-0.006	0.126	0.127
	5	20				5	100			
WG			-0.303	0.077	0.303			-0.301	0.032	0.301
WGbc			-0.005	0.095	0.095			-0.021	0.041	0.045
GMMd			-0.108	0.145	0.162			-0.028	0.069	0.070
GMMs			0.067	0.108	0.121			0.011	0.050	0.051
WGob			0.002	0.101	0.102			0.020	0.045	0.049
	10	20				10	100			
WG			-0.149	0.049	0.149			-0.147	0.022	0.147
WGbc			0.007	0.057	0.057			0.001	0.026	0.026
GMMd			-0.005	0.090	0.091			0.002	0.037	0.036
GMMs			0.007	0.078	0.079			0.005	0.034	0.034
WGob			0.017	0.058	0.061			0.023	0.025	0.031
	20	20				2	500			
WG			-0.073	0.032	0.073			-0.700	0.029	0.700
WGbc			-0.008	0.034	0.036			-0.189	0.042	0.189
GMMd			0.000	0.053	0.053			-0.004	0.085	0.085
GMMs			0.000	0.048	0.048			-0.010	0.060	0.060
WGob			0.015	0.038	0.037			-0.005	0.056	0.056
	40	20				5	500			
WG			-0.034	0.022	0.036			-0.303	0.021	0.304
WGbc			-0.013	0.022	0.023			-0.028	0.028	0.040
GMMd			0.001	0.033	0.034			-0.007	0.044	0.045
GMMs			0.002	0.032	0.032			0.000	0.032	0.032
WGob			0.016	0.024	0.026			0.018	0.030	0.035

GMMd is the first-difference GMM estimator of Arellano and Bond (1991), GMMs is the system GMM estimator of Arellano and Bover (1995) and Blundell and Bond (1998). Both use a maximum of 3 lagged instruments, with instruments stacked when $T \geq 10$. WGbc is the bias-corrected WG estimator suggested by Kiviet (1995), with GMMs used as an initial large-N consistent estimator.

Table 2: Monte Carlo comparison of alternative dynamic panel data estimators for θ when $\theta = 0.8, \mu = 1$

	T	N	MB	MAD	MAE	T	N	MB	MAD	MAE
	2	20				2	100			
WG			-0.903	0.151	0.903			-0.900	0.065	0.900
WGbc			-0.370	0.185	0.377			-0.304	0.099	0.304
GMMd			-0.263	0.696	0.782			-0.015	0.349	0.354
GMMs			-0.044	0.286	0.282			-0.020	0.145	0.143
WGob			-0.051	0.304	0.305			-0.008	0.129	0.130
	5	20				5	100			
WG			-0.430	0.077	0.430			-0.428	0.035	0.428
WGbc			-0.100	0.082	0.105			-0.096	0.038	0.096
GMMd			-0.239	0.194	0.267			-0.054	0.091	0.097
GMMs			-0.002	0.097	0.098			0.001	0.050	0.050
WGob			-0.011	0.094	0.096			0.005	0.043	0.043
	10	20				10	100			
WG			-0.220	0.045	0.220			-0.217	0.020	0.217
WGbc			-0.030	0.050	0.052			-0.027	0.023	0.031
GMMd			-0.012	0.108	0.109			0.002	0.044	0.044
GMMs			0.000	0.078	0.078			0.002	0.035	0.036
WGob			0.000	0.049	0.049			0.008	0.021	0.021
	20	20				2	500			
WG			-0.106	0.026	0.106			-0.902	0.030	0.902
WGbc			-0.024	0.030	0.034			-0.288	0.044	0.288
GMMd			-0.006	0.059	0.058			-0.004	0.146	0.148
GMMs			-0.001	0.049	0.049			-0.009	0.061	0.062
WGob			0.007	0.028	0.029			-0.007	0.057	0.059
	40	20				5	500			
WG			-0.048	0.017	0.048			-0.429	0.015	0.429
WGbc			-0.022	0.017	0.025			-0.097	0.017	0.097
GMMd			0.001	0.036	0.037			-0.014	0.041	0.041
GMMs			0.001	0.032	0.033			-0.002	0.023	0.023
WGob			0.009	0.017	0.019			0.005	0.019	0.019

See Table 1.

Table 3: Monte Carlo comparison of alternative dynamic panel data estimators for θ when $\theta = 0.4, \mu = 5$

	T	N	MB	MAD	MAE	T	N	MB	MAD	MAE
	2	20				2	100			
WG			-0.696	0.150	0.696			-0.703	0.067	0.703
WGbc			-0.084	0.167	0.162			-0.137	0.126	0.156
GMMd			-0.541	0.834	1.028			-0.156	0.620	0.651
GMMs			0.361	0.266	0.469			0.127	0.205	0.260
WGob			-0.049	0.313	0.312			-0.017	0.131	0.132
	5	20				5	100			
WG			-0.303	0.077	0.303			-0.301	0.032	0.301
WGbc			0.104	0.073	0.111			0.036	0.051	0.061
GMMd			-0.227	0.188	0.264			-0.064	0.109	0.110
GMMs			0.469	0.079	0.469			0.194	0.116	0.196
WGob			0.018	0.107	0.109			0.030	0.047	0.053
	10	20				10	100			
WG			-0.149	0.049	0.149			-0.147	0.022	0.147
WGbc			0.087	0.053	0.089			0.041	0.028	0.043
GMMd			-0.024	0.186	0.184			-0.005	0.080	0.079
GMMs			0.031	0.121	0.124			0.010	0.059	0.059
WGob			0.030	0.058	0.062			0.035	0.025	0.039
	20	20				2	500			
WG			-0.073	0.032	0.073			-0.700	0.029	0.700
WGbc			0.055	0.034	0.058			-0.181	0.070	0.181
GMMd			0.000	0.091	0.091			-0.022	0.308	0.314
GMMs			0.012	0.070	0.070			0.013	0.126	0.130
WGob			0.022	0.036	0.039			-0.001	0.059	0.059
	40	20				5	500			
WG			-0.034	0.022	0.036			-0.303	0.014	0.303
WGbc			0.028	0.022	0.032			-0.021	0.021	0.027
GMMd			0.000	0.052	0.052			-0.016	0.049	0.050
GMMs			0.002	0.042	0.041			0.021	0.032	0.033
WGob			0.018	0.023	0.027			0.031	0.021	0.031

See Table 1.

Table 4: Monte Carlo comparison of alternative dynamic panel data estimators for θ when $\theta = 0.8, \mu = 5$

	T	N	MB	MAD	MAE	T	N	MB	MAD	MAE
	2	20				2	100			
WG			-0.903	0.151	0.903			-0.900	0.065	0.900
WGbc			-0.298	0.131	0.306			-0.274	0.102	0.274
GMMd			-0.729	0.892	1.138			-0.292	0.814	0.876
GMMs			0.108	0.201	0.237			0.046	0.144	0.163
WGob			-0.048	0.313	0.310			-0.009	0.131	0.131
	5	20				5	100			
WG			-0.430	0.077	0.430			-0.428	0.035	0.428
WGbc			-0.069	0.069	0.077			-0.079	0.035	0.079
GMMd			-0.508	0.259	0.510			-0.246	0.183	0.262
GMMs			0.146	0.053	0.148			0.062	0.051	0.077
WGob			-0.002	0.095	0.095			0.016	0.043	0.044
	10	20				10	100			
WG			-0.220	0.045	0.220			-0.217	0.020	0.217
WGbc			-0.005	0.043	0.042			-0.012	0.021	0.021
GMMd			-0.100	0.236	0.257			-0.014	0.099	0.101
GMMs			0.018	0.097	0.102			0.007	0.048	0.049
WGob			0.016	0.046	0.050			0.024	0.020	0.028
	20	20				2	500			
WG			-0.106	0.026	0.106			-0.902	0.030	0.902
WGbc			0.015	0.027	0.030			-0.294	0.056	0.294
GMMd			-0.009	0.111	0.110			-0.034	0.402	0.404
GMMs			0.011	0.065	0.064			-0.009	0.089	0.087
WGob			0.020	0.026	0.032			-0.007	0.059	0.059
	40	20				5	500			
WG			-0.048	0.017	0.048			-0.429	0.015	0.429
WGbc			0.019	0.018	0.023			-0.092	0.018	0.092
GMMd			0.000	0.060	0.060			-0.065	0.088	0.101
GMMs			0.004	0.039	0.039			0.012	0.031	0.033
WGob			0.020	0.017	0.023			0.019	0.019	0.022

See Table 1.

6 Concluding comments

Dynamic panel data models are typically estimated using GMM. These instrumental variables estimators may lead to poor finite sample properties, i.e. serious small sample bias and/or relatively large standard deviations, in case of weak instruments. Bias-corrected WG estimators perform remarkably well in many cases, but the remaining bias may still be substantial when T is relatively small. In this article, we stick to a LS estimator but remove the individual effects from an AR(1) panel using orthogonal deviations from backward means as an alternative to sample means. This is equivalent to a LS estimator where the backward mean is added as a regressor in the original model, which is extremely easy to implement in practice. This WGob estimator 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 the standard estimators in terms of bias and dispersion in the cases where these estimators are know to fail, while not performing much worse in all other cases. In future research we plan to extend the model by adding explanatory variables and allowing for non-stationary initial conditions.

References

Alvarez, J., Arellano, M., 2003. The time series and cross-section asymptotics of dynamic panel data estimators. Econometrica 71, 1121–1159.

Anderson, T., Hsiao, C., 1981. Estimation of dynamic models with error components. Journal of the American Statistical Association 76, 598–606.

Anderson, T., Hsiao, C., 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., 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-components models. Journal of Econometrics 68, 29–51.

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

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

Bun, M., Carree, M., 2005. Bias-corrected estimation in dynamic panel data models. Journal of Business and Economic Statistics 23 (2), 200–210.

Bun, M., Kiviet, J., 2006. The effects of dynamic feedbacks on ls and mm estimator accuracy in panel data models. Journal of Econometrics (forthcoming).

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

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

Kiviet, J., 2006. Judging contending estimators by simulation: Tournaments in dynamic panel data models. In: Phillips, G., Tzavalis, E. (Eds.), The Refinement of Econometric Estimation and Test Procedures. Cambridge University Press.

Nickell, S., 1981. Biases in dynamic models with fixed effects. Econometrica 49 (6), 1417–1426.

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. By continuous substitution and using assumptions (A1)-(A3), we have from (1) and (10)

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

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

such that

$$\begin{aligned} E\left[y^2_{i,t-1}\right] &= E\left[\left(\frac{\alpha_i}{1-\theta} + \sum_{j=0}^{\infty} \theta^j \varepsilon_{i,t-j-1}\right)^2\right], \\ &= \frac{\sigma^2_\alpha}{(1-\theta)^2} + \frac{\sigma^2_\varepsilon}{1-\theta^2}, \end{aligned} \tag{A-3}$$

$$\begin{aligned} E\left[\left(\bar{y}^b_{i,t-1}\right)^2\right] &= E\left[\left(\frac{\alpha_i}{1-\theta} + \frac{1}{t}\sum_{s=1}^{t}\sum_{j=0}^{\infty} \theta^j \varepsilon_{i,s-j-1}\right)^2\right], \\ &= \frac{\sigma^2_\alpha}{(1-\theta)^2} + \frac{1}{t}\left(1 - \frac{2\theta}{t}\frac{1-\theta^t}{1-\theta^2}\right)\frac{\sigma^2_\varepsilon}{(1-\theta)^2}, \end{aligned} \tag{A-4}$$

$$\begin{aligned} E\left[y_{i,t-1}\bar{y}^b_{i,t-1}\right] &= E\left[\left(\frac{\alpha_i}{1-\theta} + \sum_{j=0}^{\infty} \theta^j \varepsilon_{i,t-j-1}\right)\left(\frac{\alpha_i}{1-\theta} + \frac{1}{t}\sum_{s=1}^{t}\sum_{j=0}^{\infty} \theta^j \varepsilon_{i,s-j-1}\right)\right], \\ &= E\left[\left(\frac{\alpha_i^2}{(1-\theta)^2} + \left(\sum_{j=0}^{\infty} \theta^j \varepsilon_{i,t-j-1}\right)\left(\frac{1}{t}\sum_{s=1}^{t}\sum_{j=0}^{\infty} \theta^j \varepsilon_{i,s-j-1}\right)\right)\right], \\ &= \frac{\sigma^2_\alpha}{(1-\theta)^2} + \frac{1}{t}\frac{1-\theta^t}{1-\theta}\frac{\sigma^2_\varepsilon}{1-\theta^2}, \end{aligned} \tag{A-5}$$

Averaging (A-3)-(A-5) over N and T, the results in (23)-(25) follow immediately.

Proof of Lemma 2. Considering (24) and (25) while letting $T \to \infty$:

$$\begin{aligned}
\operatorname*{plim}_{T\to\infty} \frac{1}{NT} \bar{y}_{-1}^{b\prime} \bar{y}_{-1}^{b} &= \frac{\sigma_\alpha^2}{(1-\theta)^2} + \frac{\sigma_\varepsilon^2}{(1-\theta)^2} \frac{1}{T} \operatorname*{plim}_{T\to\infty} \sum_{t=1}^{T} \frac{1}{t} \left(1 - \frac{2\theta}{t} \frac{1-\theta^t}{1-\theta^2}\right), \\
&= \frac{\sigma_\alpha^2}{(1-\theta)^2} + \frac{\sigma_\varepsilon^2}{(1-\theta)^2} \frac{1}{T} \left(\gamma + \ln(T) - \frac{2\theta}{1-\theta^2}\left(\frac{\Pi^2}{6} - \mathrm{Li}_2(\theta)\right)\right), \\
&= \frac{\sigma_\alpha^2}{(1-\theta)^2}, \qquad \text{(A-6)} \\
\operatorname*{plim}_{T\to\infty} \frac{1}{NT} \bar{y}_{-1}^{b\prime} y_{-1} &= \frac{\sigma_\alpha^2}{(1-\theta)^2} + \frac{\sigma_\varepsilon^2}{1-\theta^2} \frac{1}{T} \operatorname*{plim}_{T\to\infty} \sum_{t=1}^{T} \frac{1}{t} \frac{1-\theta^t}{1-\theta}, \\
&= \frac{\sigma_\alpha^2}{(1-\theta)^2} + \frac{\sigma_\varepsilon^2}{1-\theta^2} \frac{1}{1-\theta} \frac{1}{T} (\gamma + \ln(T) + \ln(1-\theta)), \\
&= \frac{\sigma_\alpha^2}{(1-\theta)^2}, \qquad \text{(A-7)}
\end{aligned}$$

where use is made of

$$\lim_{T\to\infty} \sum_{t=1}^{T} \frac{1}{t} = \gamma + \ln(T), \qquad \lim_{T\to\infty} \sum_{t=1}^{T} \frac{\theta^t}{t} = -\ln(1-\theta), \qquad \lim_{T\to\infty} \sum_{t=1}^{T} \frac{1}{t^2} = \frac{\pi^2}{6}, \qquad \text{(A-8)}$$

with γ being the Euler-Mascheroni constant $0.57721\ldots$. The dilogarithm $\lim_{T\to\infty} \sum_{t=1}^{T} \theta^t / t^2 \equiv \mathrm{Li}_2(\theta)$ cannot be evaluated in closed form for all values of θ but is convergent for the relevant range $-1 \leqslant \theta \leqslant 1$. The result in (26) follows from dividing (A-7) by (A-6).

Using (A-1), (A-2), assumption (A1) and $Q^b D = 0$

$$\begin{aligned}
E\left[\widetilde{y}_{i,t-1}^{b} (D\alpha + \varepsilon_{it})\right] &= E\left[\left(y_{i,t-1} - \bar{y}_{it,-1}^{b}\right) \varepsilon_{it}\right] \\
&= -E\left[\left(\sum_{j=0}^{\infty} \theta^j \varepsilon_{i,t-j-1} - \frac{1}{t} \sum_{s=1}^{t} \sum_{j=0}^{\infty} \theta^j \varepsilon_{i,s-j-1}\right) \varepsilon_{it}\right] \\
&= 0, \qquad \text{(A-9)}
\end{aligned}$$

such that averaging over N and T yields the result $\frac{1}{NT} E\left[\widetilde{y}_{-1}^{b} \varepsilon\right] = 0$ in (28). Using (23)-(25) it follows immediately

$$\begin{aligned}
\frac{1}{NT} E\left[\widetilde{y}_{-1}^{b\prime} y_{-1}\right] &= \frac{1}{NT} E\left[\left(y_{-1} - \bar{y}_{-1}^{b}\right)' y_{-1}\right], \\
&= \sigma_{y_{-1}}^2 - \sigma_{y_{-1}\bar{y}_{-1}^{b}}, \\
&= \frac{\sigma_\varepsilon^2}{1-\theta^2} - \frac{\sigma_\varepsilon^2}{1-\theta^2} \frac{1}{T} \sum_{t=1}^{T} \frac{1}{t} \frac{1-\theta^t}{1-\theta} \\
&= \frac{\sigma_\varepsilon^2}{1-\theta^2} \left(1 - \frac{1}{T} \sum_{t=1}^{T} \frac{1}{t} \frac{1-\theta^t}{1-\theta}\right), \qquad \text{(A-10)}
\end{aligned}$$

which is the result in (29).

Under assumptions (A1), (A2) and (A3) and letting $N \to \infty$, the results in (30) and (31) are obtained as:

$$
\begin{aligned}
\operatorname*{plim}_{N\to\infty} \frac{1}{NT} \widehat{y}^{b\prime}_{-1} (D\alpha + \varepsilon) &= \operatorname*{plim}_{N\to\infty} \frac{1}{NT} \left(Q^b_{\perp} y_{-1}\right)' (D\alpha + \varepsilon), \\
&= \operatorname*{plim}_{N\to\infty} \frac{1}{NT} y'_{-1} \left(I - \bar{y}^b \left(\bar{y}^{b\prime}\bar{y}^b\right)^{-1} \bar{y}^{b\prime}\right) (D\alpha + \varepsilon), \\
&= \operatorname*{plim}_{N\to\infty} \frac{1}{NT} \left(\frac{\alpha' D'}{1-\theta} - y'_{-1}\bar{y}^b \left(\bar{y}^{b\prime}\bar{y}^b\right)^{-1} \frac{\alpha' D'}{1-\theta}\right) D\alpha, \\
&= \operatorname*{plim}_{N\to\infty} \frac{1}{NT} \left(1 - y'_{-1}\bar{y}^b \left(\bar{y}^{b\prime}\bar{y}^b\right)^{-1}\right) \frac{\alpha' D' D\alpha}{1-\theta}, \\
&= \left(1 - \frac{\sigma_{y_{-1}\bar{y}^b_{-1}}}{\sigma^2_{\bar{y}^b_{-1}}}\right) \frac{\sigma^2_\alpha}{1-\theta}, \\
&= \left(1 - \delta_{y_{-1}\bar{y}^b_{-1}}\right) \frac{\sigma^2_\alpha}{1-\theta},
\end{aligned} \tag{A-11}
$$

where $\delta_{y_{-1}\bar{y}^b_{-1}} = \sigma_{y_{-1}\bar{y}^b_{-1}} \Big/ \sigma^2_{\bar{y}^b_{-1}}$ and

$$
\begin{aligned}
\operatorname*{plim}_{N\to\infty} \frac{1}{NT} \widehat{y}^{b\prime}_{-1}\widehat{y}^b_{-1} &= \operatorname*{plim}_{N\to\infty} \frac{1}{NT} y'_{-1} Q^b_{\perp} y_{-1}, \\
&= \operatorname*{plim}_{N\to\infty} \frac{1}{NT} y'_{-1} \left(I - \bar{y}^b \left(\bar{y}^{b\prime}\bar{y}^b\right)^{-1} \bar{y}^{b\prime}\right) y_{-1}, \\
&= \sigma^2_{y_{-1}} - \delta_{y_{-1}\bar{y}^b_{-1}} \sigma_{y_{-1}\bar{y}^b_{-1}}, \\
&= \left(1 - \rho^2_{y_{-1}\bar{y}^b_{-1}}\right) \sigma^2_{y_{-1}},
\end{aligned} \tag{A-12}
$$

where $\rho_{y_{-1}\bar{y}^b_{-1}} = \sigma_{y_{-1}\bar{y}^b_{-1}} \Big/ \left(\sigma_{y_{-1}} \sigma_{\bar{y}^b_{-1}}\right)$.

Proof of Theorem 1. The consistency of $\widehat{\theta}^{WG}_{\perp}$ for $T \to \infty$ follows directly from (28) which implies that $\operatorname*{plim}_{T\to\infty} \widetilde{y}^{b\prime}_{-1}\varepsilon / NT = 0$, and (29) which implies that $\operatorname*{plim}_{T\to\infty} \widetilde{y}^{b\prime}_{-1} y_{-1} / NT = \dfrac{\sigma^2_\varepsilon}{1-\theta^2}$, where the latter follows from (A-8).

The inconsistency for $N \to \infty$ of the $\widehat{\theta}^{WG}_{\perp}$ is given by

$$
\begin{aligned}
\operatorname*{plim}_{N\to\infty} \left(\widehat{\theta}^{WG}_{\perp} - \theta\right) &= \frac{\operatorname*{plim}_{N\to\infty} \widehat{y}^{b\prime}_{-1} (D\alpha + \varepsilon)}{\operatorname*{plim}_{N\to\infty} \widehat{y}^{b\prime}_{-1}\widehat{y}^b_{-1}}, \\
&= \frac{1 - \delta_{y_{-1}\bar{y}^b_{-1}}}{\sigma^2_{y_{-1}} - \delta_{y_{-1}\bar{y}^b_{-1}} \sigma_{y_{-1}\bar{y}^b_{-1}}} \frac{\sigma^2_\alpha}{1-\theta}, && \text{(A-13)} \\
&= \frac{\left(1 - \delta_{y_{-1}\bar{y}^b_{-1}}\right)}{\left(1 - \rho^2_{y_{-1}\bar{y}^b_{-1}}\right) \sigma^2_{y_{-1}}} \frac{\sigma^2_\alpha}{1-\theta}. && \text{(A-14)}
\end{aligned}
$$

Inserting (23) and (25) in (A-13):

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

where use is made of

$$\delta_{y_{-1}\bar{y}^b_{-1}} = \frac{1+\frac{1}{T}\sum_{t=1}^T\frac{1}{t}\frac{1-\theta^t}{1+\theta}\frac{\sigma_\varepsilon^2}{\sigma_\alpha^2}}{1+\frac{1}{T}\sum_{t=1}^T\frac{1}{t}\left(1-\frac{2\theta}{t}\frac{1-\theta^t}{1-\theta^2}\right)\frac{\sigma_\varepsilon^2}{\sigma_\alpha^2}}. \tag{A-15}$$

Proof of Corollary 1.

(a) From (32) it follows that for $-1<\theta<1$ and $T\geq 3$ $\operatorname*{plim}_{N\to\infty}\left(\widehat{\theta}_{\perp}^{WG}-\theta\right)>0$ when $\delta_{y_{-1}\bar{y}^b_{-1}}<1$. From the definition of $\delta_{y_{-1}\bar{y}^b_{-1}}$ in (A-15) we have that $\delta_{y_{-1}\bar{y}^b_{-1}}<1$ if

$$\begin{aligned}
\sum_{t=1}^T\frac{1}{t}\frac{1-\theta^t}{1+\theta} &< \sum_{t=1}^T\frac{1}{t}\left(1-\frac{2\theta}{t}\frac{1-\theta^t}{1-\theta^2}\right),\\
\frac{1-\theta^t}{1+\theta} &< 1-\frac{2\theta}{t}\frac{1-\theta^t}{1-\theta^2} \quad \forall t\geq 3,\\
1-\theta^t &< 1-\theta^t+\theta\frac{t-2}{t}\left(1-\frac{2}{t-2}\theta-\ldots-\frac{2}{t-2}\theta^{t-1}+\theta^t\right) \quad \forall t\geq 3.
\end{aligned}$$

As for $-1<\theta<1$ and $t\geq 3$ the term $\left(1-\frac{2}{t-2}\theta-\ldots-\frac{2}{t-2}\theta^{t-1}+\theta^t\right)>0$, $\delta_{y_{-1}\bar{y}^b_{-1}}<1$ when $\theta>0$. When $\theta<0$, $\delta_{y_{-1}\bar{y}^b_{-1}}>1$ such that $\operatorname*{plim}_{N\to\infty}\left(\widehat{\theta}_{\perp}^{WG}-\theta\right)<0$.

(b) The upper bound in (37) is obtained by setting $\frac{\sigma_\varepsilon^2}{\sigma_\alpha^2}=0$ in (33).

(c) Follows directly from (33), where for the case $\theta=1$ use is made of $\frac{1-\theta^t}{1-\theta}=t$.

(d) For $T=2$, (37) is given by

$$\underset{N,\frac{\sigma_{\varepsilon}^{2}}{\sigma_{\alpha}^{2}}\to 0}{\text{plim}}\left(\widehat{\theta}_{\perp}^{WG}-\theta\right)=\frac{\theta\left(1-\theta\right)\left(0+0\right)}{2\left(1-\theta\right)-\frac{1}{2}\left(1-\theta\right)\left(3+\theta\right)}=0, \tag{A-16}$$

while for $T=3$ we have

$$\begin{aligned}\underset{N,\frac{\sigma_{\varepsilon}^{2}}{\sigma_{\alpha}^{2}}\to 0}{\text{plim}}\left(\widehat{\theta}_{\perp}^{WG}-\theta\right)&=\frac{\theta\left(1-\theta\right)\left(0+0+\frac{1}{3}\left(1+\theta^{2}-\frac{2}{3}\left(1+\theta+\theta^{2}\right)\right)\right)}{3\left(1-\theta\right)+\left(-\frac{1}{2}\left(1-\theta\right)\left(3+\theta\right)+\frac{1}{3}\left(2\theta^{3}-\left(1-\theta\right)-\frac{2\theta}{3}\left(1+\theta+\theta^{2}\right)\right)\right)},\\
&=\frac{\theta\left(1-\theta\right)\frac{1}{9}\left(1-\theta\right)^{2}}{3\left(1-\theta\right)-\frac{1}{9}\frac{1}{2}\left(1-\theta\right)\left(11\left(3+\theta\right)+2\theta\left(1+4\theta\right)\right)},\\
&=\frac{\theta\left(1-\theta\right)}{\frac{1}{2}\left(21+8\theta\right)}=\frac{\theta\left(1-\theta\right)}{4\left(3-3/8+\theta\right)}.\end{aligned} \tag{A-17}$$

In order to derive an approximation for large T, first note that

$$\begin{aligned}A_T&=\frac{1}{T}\sum_{t=1}^{T}\frac{1}{t}\left(1+\theta^{t-1}-\frac{2}{t}\frac{1-\theta^{t}}{1-\theta}\right)\\
&=\left(1-\theta\right)^{2}\frac{1}{T}\sum_{t=3}^{T}\frac{1}{t}\left(\frac{1}{t}\sum_{j=1}^{t-2}\left(j\left(t-j-1\right)\right)\theta^{j-1}\right)\\
&=\left(1-\theta\right)^{2}\frac{1}{T}\sum_{t=2}^{T}\frac{1}{t}\left(\frac{1-\theta^{t-2}}{1-\theta}+\frac{1}{t}\sum_{j=1}^{t-2}\left(t\left(j-1\right)-j\left(1+j\right)\right)\theta^{j-1}\right)\\
&\simeq\left(1-\theta\right)^{2}\frac{1}{T}\sum_{t=2}^{T}\frac{1}{t}\frac{1-\theta^{t-2}}{1-\theta}=\left(1-\theta\right)\frac{1}{T}\sum_{t=2}^{T}\frac{1-\theta^{t-2}}{t}=\left(1-\theta\right)\frac{\ln\left(T\right)}{T}\end{aligned} \tag{A-18}$$

Inserting (A-18) in (33) yields

$$\underset{N\to\infty}{\text{plim}}\left(\widehat{\theta}_{\perp}^{WG}-\theta\right)\simeq\frac{\theta\left(1-\theta\right)\ln\left(T\right)}{T} \tag{A-19}$$

where use is made of

$$\begin{aligned}B_T&\simeq\frac{1}{T}\left(-2\ln\left(1-\theta\right)-\left(1-\theta\right)\left(\gamma+\ln\left(T\right)\right)-\frac{2\theta}{1-\theta}\left(\frac{\pi^{2}}{6}-\text{Li}_{2}\left(\theta\right)\right)\right),\\
&=-\left(1-\theta\right)\frac{\ln\left(T\right)}{T}+O\left(T^{-1}\right),\end{aligned} \tag{A-20}$$

$$\begin{aligned}C_T&\simeq\frac{1}{T}\left(\left(1-\theta\right)\left(\gamma+\ln\left(T\right)\right)-\frac{2\theta}{1+\theta}\left(\frac{\pi^{2}}{6}-\text{Li}_{2}\left(\theta\right)\right)-\frac{1}{1+\theta}\frac{1}{T}\left(\frac{\pi^{2}}{6}-2\text{Li}_{2}\left(\theta\right)+2\text{Li}_{2}\left(\theta^{2}\right)\right)\right),\\
&=\left(1-\theta\right)\frac{\ln\left(T\right)}{T}+O\left(T^{-1}\right).\end{aligned} \tag{A-21}$$

(f) The approximation in (37) shows that the inconsistency of $\widehat{\theta}_{\perp}^{WG}$ is $O\left(\ln\left(T\right)/T\right)$.