UvA ECONOMETRICS

Discussion Paper: 2007/01

The weak instrument problem of the system GMM estimator in dynamic panel data models

Maurice J.G. Bun and Frank Windmeijer

(revised July 2009)

www.fee.uva.nl/ke/UvA-Econometrics

Amsterdam School of Economics
Department of Quantitative Economics
Roetersstraat 11
1018 WB AMSTERDAM
The Netherlands

UvA UNIVERSITEIT VAN AMSTERDAM

The Weak Instrument Problem of the System GMM Estimator in Dynamic Panel Data Models

Maurice J.G. Bun
Dept. of Quantitative Economics
University of Amsterdam
M.J.G.Bun@uva.nl

Frank Windmeijer
Dept. of Economics
University of Bristol
and cemmap/IFS
f.windmeijer@bristol.ac.uk

July, 2009

Abstract

The system GMM estimator for dynamic panel data models combines moment conditions for the model in first differences with moment conditions for the model in levels. It has been shown to improve on the GMM estimator in the first differenced model in terms of bias and root mean squared error. However, we show in this paper that in the covariance stationary panel data AR(1) model the expected values of the concentration parameters in the differenced and levels equations for the cross-section at time t are the same when the variances of the individual heterogeneity and idiosyncratic errors are the same. This indicates a weak instrument problem also for the equation in levels. We show that the 2SLS biases relative to that of the OLS biases are then similar for the equations in differences and levels, as are the size distortions of the Wald tests. These results are shown to extend to the panel data GMM estimators.

JEL Classification: C12, C13, C23
Keywords: Dynamic Panel Data, System GMM, Weak Instruments

1 Introduction

A commonly employed estimation procedure to estimate the parameters in a dynamic panel data model with unobserved individual specific heterogeneity is to transform the model into first differences. Sequential moment conditions are then used where lagged levels of the variables are instruments for the endogenous differences and the parameters estimated by GMM, see Arellano and Bond (1991). It has been well documented (see e.g. Blundell and Bond (1998)) that this GMM estimator in the first differenced (DIF) model can have very poor finite sample properties in terms of bias and precision when the series are persistent, as the instruments are then weak predictors of the endogenous changes. Blundell and Bond (1998) proposed the use of extra moment conditions that rely on certain stationarity conditions of the initial observation, as suggested by Arellano and Bover (1995). When these conditions are satisfied, the resulting system (SYS) GMM estimator has been shown in Monte Carlo studies by e.g. Blundell and Bond (1998) and Blundell, Bond and Windmeijer (2000) to have much better finite sample properties in terms of bias and root mean squared error (rmse) than that of the DIF GMM estimator.

The additional moment conditions of the SYS estimator can be shown to correspond to the model in levels (LEV), with lagged differences of the endogenous variables as instruments. Blundell and Bond (1998) argued that the SYS GMM estimator performs better than the DIF GMM estimator because the instruments in the LEV model remain good predictors for the endogenous variables in this model even when the series are very persistent. They showed for an AR(1) panel data model that the reduced form parameters in the LEV model do not approach 0 when the autoregressive parameter approaches 1, whereas the reduced form parameters in the DIF model do.

Because of the good performance of the SYS GMM estimator relative to the DIF GMM estimator in terms of finite sample bias and rmse, it has become the estimator of choice in many applied panel data settings. Among the many examples where the

SYS GMM estimator has been used are the estimation of production functions and technological spillovers using firm level panel data (see e.g. Levinsohn and Petrin (2003) and Griffith, Harrison and Van Reenen (2006)), the estimation of demand for addictive goods using consumer level panel data (see e.g. Picone, Sloan and Trogdon (2004)) and the estimation of growth models using country level panel data (see e.g. Levine, Loayza and Beck (2000) and Bond, Hoeffler and Temple (2001)). The country level panel data in particular are characterised by highly persistent series (e.g. output or financial data) and a relatively small number of countries and time periods. The variance of the country effects is furthermore often expected to be quite high relative to the variance of the transitory shocks. As we show here, these characteristics combined may lead to a weak instrument problem also for the SYS GMM estimator.

For a simple cross-section linear IV model, a measure of the information content of the instruments is the so-called concentration parameter (see e.g. Rothenberg (1984)). In this paper we calculate the expected concentration parameters for the LEV and DIF reduced form models in a covariance stationary AR(1) panel data model. We do this per time period, i.e. we consider the estimation of the parameter using the moment conditions for a single cross-section only for any given time period. We show that the expected concentration parameters are equal in the LEV and DIF models when the variance of the unobserved heterogeneity term that is constant over time (σ_η^2) is equal to the variance of the idiosyncratic shocks (σ_v^2). This is exactly the environment under which most Monte Carlo results were obtained that showed the superiority of the SYS GMM estimator relative to the DIF GMM estimator. However, the equality in expectation of the concentration parameters indicates that there is also a weak instrument problem in the LEV model when the series are persistent.

If the expected concentration parameters are the same, why is it that the extra information from the LEV moment conditions results in an estimator that has such superior finite sample properties in terms of bias and rmse? We first of all show that the bias of the OLS estimators in the DIF and LEV structural models are very different. The (absolute) bias of the LEV OLS estimator is much smaller than that of the OLS estimator in the DIF model when the series are very persistent. Using the results of higher order expansions, we argue and show in Monte Carlo simulations that the biases of the LEV and DIF cross-sectional 2SLS estimators, relative to the biases of their respective OLS estimators, are the same when $\sigma_\eta^2 = \sigma_v^2$. Therefore the absolute bias of the LEV 2SLS estimator is smaller than that of the DIF 2SLS estimator when the series are persistent.

Further expansion results as in Morimune (1989) indicate that we can expect the size distortions of the Wald tests to be similar in the cross-sectional 2SLS DIF and LEV models when the expected concentration parameters are the same. This is confirmed by a Monte Carlo analysis. When the expected concentration parameters are small, which happens when the series are very persistent, the size distortions of the Wald tests can become substantial. As the SYS 2SLS estimator is a weighted average of the DIF and LEV 2SLS estimators, with the weight on the LEV moment conditions increasing with increasing persistence of the series, the results for the SYS estimator mimic that of the LEV estimator quite closely.

The expectation of the LEV concentration parameter is larger than that of the DIF model when σ_η^2 is smaller than σ_v^2, and the relative biases of LEV and SYS 2SLS estimators are smaller and the associated Wald tests perform better than those of DIF. The reverse is the case when σ_η^2 is larger than σ_v^2. Also, unlike for DIF, the LEV OLS bias increases with increasing variance ratio, $vr = \sigma_\eta^2/\sigma_v^2$, and therefore the performances of the LEV and SYS 2SLS estimators deteriorate with increasing vr. These results are shown to extend to the panel data setting when estimating the model by GMM and are in line with the finite sample bias approximation results of Bun and Kiviet (2006) and Hayakawa (2007), and with the findings from an extensive Monte Carlo study by Kiviet

(2007). Furthermore, our theoretical results provide a rationale for the poor performance of the SYS GMM Wald test when data are persistent, as found by Bond and Windmeijer (2005).

For the covariance stationary AR(1) panel data model our results therefore show that the SYS GMM estimator has indeed a smaller bias and rmse than DIF GMM when the series are persistent, but that this bias increases with increasing $vr = \sigma_\eta^2/\sigma_v^2$ and can become substantial. The Wald test can be severely size distorted for both DIF and SYS GMM with persistent data, but the SYS Wald test size properties deteriorate further with increasing vr. These results follow from the weak instrument problem that is also present in the LEV moment conditions.

The setup of the paper is as follows. Section 2 introduces the AR(1) panel data model, the moment conditions and GMM estimators. Section 3 briefly discusses the concentration parameter in a simple cross-section setting. Section 4 calculates the expected concentration parameters for the DIF and LEF models for cross-section analysis of the AR(1) panel data model, presents the OLS biases and some Monte Carlo and theoretical results on (relative) biases and Wald tests size distortions for the 2SLS estimators. Section 5 presents Monte Carlo and some analytical results for the GMM panel data estimators. Section 6 concludes.

2 Model and GMM Estimators

We consider the first-order autoregressive panel data model

$$\begin{aligned} y_{it} &= \alpha y_{i,t-1} + u_{it}, \qquad i = 1, ..., n;\ t = 2, ..., T, \qquad (1) \\ u_{it} &= \eta_i + v_{it} \end{aligned}$$

where it is assumed that η_i and v_{it} have an error components structure with

$$E(\eta_i) = 0,\ E(v_{it}) = 0,\ E(v_{it}\eta_i) = 0, \qquad i = 1, ..., n;\ t = 2, ..., T \qquad (2)$$

$$E\left(v_{it}v_{is}\right)=0, \qquad i=1,...,n \text{ and } t\neq s, \tag{3}$$

and the initial condition satisfies

$$E\left(y_{i1}v_{it}\right)=0, \qquad i=1,...,n\ t=2,...,T. \tag{4}$$

Under these assumptions the following $(T-1)(T-2)/2$ linear moment conditions are valid

$$E\left(y_i^{t-2}\Delta u_{it}\right)=0, \qquad t=3,...,T, \tag{5}$$

where $y_i^{t-2}=(y_{i1},y_{i2},...,y_{it-2})'$ and $\Delta u_{it}=u_{it}-u_{i,t-1}=\Delta y_{it}-\alpha\Delta y_{i,t-1}$.

Defining

$$Z_{di}=\begin{bmatrix} y_{i1} & 0 & 0 & \cdots & 0 & \cdots & 0 \\ 0 & y_{i1} & y_{i2} & \cdots & 0 & \cdots & 0 \\ . & . & . & \cdots & . & \cdots & . \\ 0 & 0 & 0 & \cdots & y_{i1} & \cdots & y_{iT-2} \end{bmatrix};\ \Delta u_i=\begin{bmatrix} \Delta u_{i3} \\ \Delta u_{i4} \\ \vdots \\ \Delta u_{iT} \end{bmatrix},$$

moment conditions (5) can be more compactly written as

$$E\left(Z'_{di}\Delta u_i\right)=0, \tag{6}$$

and the GMM estimator for α is given by (see e.g. Arellano and Bond (1991))

$$\widehat{\alpha}_d=\frac{\Delta y'_{-1}Z_dW_n^{-1}Z'_d\Delta y}{\Delta y'_{-1}Z_dW_n^{-1}Z'_d\Delta y_{-1}}$$

where $\Delta y=(\Delta y'_1,\Delta y'_2...\Delta y'_n)'$, $\Delta y_i=(\Delta y_{i3},\Delta y_{i4},...,\Delta y_{iT})'$, Δy_{-1} the lagged version of Δy, $Z_d=(Z'_{d1},Z'_{d2},...,Z'_{dn})'$ and W_n is a weight matrix determining the efficiency properties of the GMM estimator. Clearly, $\widehat{\alpha}_d$ is a GMM estimator in the differenced model and we refer to it as the DIF-GMM estimator, and moment conditions (5) or (6) as the DIF moment conditions.

Blundell and Bond (1998) exploit additional moment conditions from the assumption on the initial condition (see Arellano and Bover (1995)) that

$$E\left(\eta_i\Delta y_{i2}\right)=0, \tag{7}$$

which holds when the process is mean stationary:

$$y_{i1} = \frac{\eta_i}{1-\alpha} + \varepsilon_i, \tag{8}$$

with $E(\varepsilon_i) = E(\varepsilon_i \eta_i) = 0$. If (2), (3), (4) and (7) hold then the following $(T-1)(T-2)/2$ moment conditions are valid

$$E\left(u_{it} \Delta y_i^{t-1}\right) = 0, \qquad t = 3, ..., T, \tag{9}$$

where $\Delta y_i^{t-1} = (\Delta y_{i2}, \Delta y_{i3}, ..., \Delta y_{it-1})'$. Defining

$$Z_{li} = \begin{bmatrix} \Delta y_{i2} & 0 & 0 & \cdots & 0 & \cdots & 0 \\ 0 & \Delta y_{i2} & \Delta y_{i3} & \cdots & 0 & \cdots & 0 \\ . & . & . & \cdots & . & \cdots & . \\ 0 & 0 & 0 & \cdots & \Delta y_{i2} & \cdots & \Delta y_{iT-1} \end{bmatrix}; \; u_i = \begin{bmatrix} u_{i3} \\ u_{i4} \\ \vdots \\ u_{iT} \end{bmatrix},$$

moment conditions (9) can be written as

$$E\left(Z'_{li} u_i\right) = 0, \tag{10}$$

with the GMM estimator based on these moment conditions given by

$$\widehat{\alpha}_l = \frac{y'_{-1} Z_l W_n^{-1} Z'_l y}{y'_{-1} Z_l W_n^{-1} Z'_l y_{-1}},$$

where we will refer to $\widehat{\alpha}_l$ as the LEV-GMM estimator, and (9) or (10) as the LEV moment conditions.

The full set of linear moment conditions under assumptions (2), (3), (4) and (7) is given by

$$\begin{aligned} E\left(y_i^{t-2} \Delta u_{it}\right) &= 0 \qquad t = 3, ..., T; \\ E\left(u_{it} \Delta y_{i,t-1}\right) &= 0 \qquad t = 3, ..., T, \end{aligned} \tag{11}$$

or

$$E\left(Z'_{si} p_i\right) = 0, \tag{12}$$

where

$$Z_{si} = \begin{bmatrix} Z_{di} & 0 & \cdots & 0 \\ 0 & \Delta y_{i2} & & 0 \\ . & . & \ddots & . \\ 0 & 0 & \cdots & \Delta y_{iT} \end{bmatrix}; \; p_i = \begin{bmatrix} \Delta u_i \\ u_i \end{bmatrix}.$$

The GMM estimator based on these moment conditions is

$$\widehat{\alpha}_s = \frac{q'_{-1} Z_s W_n^{-1} Z'_s q}{q'_{-1} Z_s W_n^{-1} Z'_s q_{-1}}$$

with $q_i = (\Delta y'_i, y'_i)'$. This estimator is called the system or SYS-GMM estimator, see Blundell and Bond (1998), and we refer to moment conditions (11) or (12) as the SYS moment conditions.

In most derivations below, we further assume that the initial observation is drawn from the covariance stationary distribution, implying that $E\left(\varepsilon_i^2\right) = \frac{\sigma_v^2}{1-\alpha^2}$ in (8).

3 Concentration Parameter

Consider the simple linear cross section model with one endogenous regressor x and k_z instruments z

$$\begin{aligned} y_i &= x_i \beta + u_i \\ x_i &= z'_i \pi + \xi_i, \end{aligned} \tag{13}$$

for $i = 1, ..., n$, where the (u_i, ξ_i) are independent draws from a bivariate normal distribution with zero means, variances σ_u^2 and σ_ξ^2, and correlation coefficient ρ. The parameter β is estimated by 2SLS:

$$\widehat{\beta} = \frac{x' P_Z y}{x' P_Z x},$$

where $P_Z = Z\left(Z'Z\right)^{-1} Z'$.

It is well known that when instruments are weak, i.e. when they are only weakly correlated with the endogenous regressor, the 2SLS estimator can perform poorly in finite samples, see e.g. Bound, Jaeger and Baker (1995), Staiger and Stock (1997), Stock, Wright and Yogo (2002) and Stock and Yogo (2005). With weak instruments, the 2SLS estimator is biased in the direction of the OLS estimator, and its distribution is non-normal which affects inference using the Wald testing procedure.

A measure of the strength of the instruments is the concentration parameter, which is defined as

$$\mu = \frac{\pi' Z' Z \pi}{\sigma_\xi^2}.$$

When it is evaluated at the OLS, first stage, estimated parameters

$$\widehat{\mu} = \frac{\widehat{\pi}' Z' Z \widehat{\pi}}{\widehat{\sigma}_\xi^2},$$

it is clear that $\widehat{\mu}$ is equal to the Wald test for testing the hypothesis $H_0 : \pi = 0$, and $\widehat{\mu}/k_z$ equal to the F-test statistic. Bound, Jaeger and Baker (1995) and Staiger and Stock (1997) advocate use of the first-stage F-test to investigate the strength of the instruments.

Rothenberg (1984) shows how the concentration parameter relates to the distribution of the IV estimator by means of the following expansion

$$\widehat{\beta} = \beta + \frac{\pi' Z' u + \xi' P_Z u}{\pi' Z' Z \pi + 2\pi' Z' \xi + \xi' P_Z \xi}, \tag{14}$$

and so

$$\sqrt{\mu}\left(\widehat{\beta} - \beta\right) = \frac{\sigma_u}{\sigma_\xi} \frac{A + \frac{s}{\sqrt{\mu}}}{1 + 2\left(\frac{B}{\sqrt{\mu}}\right) + \frac{S}{\mu}},$$

where

$$A = \frac{\pi' Z' u}{\sigma_u \sqrt{\pi' Z' Z \pi}}; \quad B = \frac{\pi' Z' \xi}{\sigma_\xi \sqrt{\pi' Z' Z \pi}}$$

$$s = \frac{\xi' P_Z u}{\sigma_\xi \sigma_u}; \quad S = \frac{\xi' P \xi}{\sigma_\xi^2}.$$

(A, B) is bivariate normal with zero means, unit variances and correlation coefficient ρ. s has mean $k_z \rho$ and variance $k_z\left(1+\rho^2\right)$ and S has mean k_z and variance $2k_z$. It is clear that when μ is large, $\sqrt{\mu}\left(\widehat{\beta} - \beta\right)$ behaves like the $n\left(0, \sigma_u^2/\sigma_\xi^2\right)$ random variable.

The concentration parameter μ is a key quantity in describing the finite sample properties of the IV estimator. The approximate bias of the 2SLS estimator can be obtained using higher order asymptotics based on the expansion in (14), see Nagar (1959) , Buse (1992) and Hahn and Kuersteiner (2002). Following Hahn and Kuersteiner (2002), the bias is derived from the expansion

$$E\left(n^{1/2}\left(\widehat{\beta}_{2SLS}-\beta\right)\right) \approx E\left(\frac{\pi' z_u}{\pi' Q\pi}\right)+n^{-1/2}\left(E\left(\frac{z_\xi' Q z_u}{\pi' Q\pi}\right)-2E\left(\frac{(\pi' z_\xi)(\pi' z_u)}{(\pi' Q\pi)^2}\right)\right) \quad (15)$$

where $z_u = \frac{1}{\sqrt{n}}Z'u$, $z_\xi = \frac{1}{\sqrt{n}}Z'\xi$ and $Q = E\left(\frac{1}{n}z_i z_i'\right)$. It follows that the approximate bias of the IV estimator can be expressed as

$$E\left(\widehat{\beta}_{2SLS}\right)-\beta \approx \frac{1}{n}\frac{(k_z-2)\,\sigma_{u\xi}}{\pi' Q\pi} = \frac{\sigma_{u\xi}}{\sigma_\xi^2}\frac{(k_z-2)}{nE\left(\frac{1}{n}\mu\right)}. \quad (16)$$

Hence the bias is inversely proportional to the value of the concentration parameter. It does not only depend on the concentration parameter, but also on the number of instruments k_z and the degree of endogeneity embodied in the covariance $\sigma_{u\xi}$. However, the relevance of the concentration parameter for finite sample bias becomes even more pronounced when we consider the absolute bias of the IV estimator, relative to that of the OLS estimator as defined by

$$\text{RelBias} = \frac{\left|E\left(\widehat{\beta}_{2SLS}\right)-\beta\right|}{\left|E\left(\widehat{\beta}_{OLS}\right)-\beta\right|},$$

see e.g. Bound et al. (1995). The bias of the OLS estimator can be approximated by (see e.g. Hahn and Hausman (2002))

$$E\left(\widehat{\beta}_{OLS}\right)-\beta \approx \frac{\sigma_{u\xi}}{\pi' Q\pi+\sigma_\xi^2} = \frac{\sigma_{u\xi}}{\sigma_\xi^2}\frac{1}{E\left(\frac{1}{n}\mu\right)+1},$$

which is equal to inconsistency of OLS. The relative bias is then approximately given by

$$\text{RelBias} \approx \frac{(k_z-2)\left(E\left(\frac{1}{n}\mu\right)+1\right)}{nE\left(\frac{1}{n}\mu\right)}, \quad (17)$$

i.e. a function of $E\left(\frac{1}{n}\mu\right)$, n, and k_z only.

The concentration parameter is further an important element in describing size distortions of t or Wald tests based on the 2SLS estimator. For μ large the standard 2SLS

t-ratio for testing $H_0 : \beta = \beta_0$ behaves approximately as standard normal. Morimune (1989) derives a higher-order expansion of this conventional 2SLS t-ratio. Applying Theorem 2 of Morimune (1989) we find for the set-up with one endogenous regressor and no additional exogenous regressors that the $O(n^{-1/2})$ and $O(n^{-1})$ terms in the expansion of the 2SLS t-statistic only depend on μ, k_z and $\rho_{u\xi}$. The latter quantity is the correlation coefficient of u and ξ. Moreover, for a two-sided t-test the $O(n^{-1/2})$ term cancels in the approximation.

All results discussed above are based on conventional higher-order asymptotics, i.e. assuming strong identification. Hence, these higher-order approximations may not always be informative in case of weak instruments. However, regarding the relevance of the concentration parameter, weak instrument asymptotics as derived by Staiger and Stock (1997) lead to similar conclusions compared with conventional fixed-parameter higher-order asymptotics. Staiger and Stock (1997) develop weak instrument asymptotics by setting $\pi = \pi_n = C/\sqrt{n}$, in which case the concentration parameter converges to a constant. They then show that 2SLS is not consistent and has a nonstandard asymptotic distribution. These results are of course different from conventional asymptotics. However, Staiger and Stock (1997) show that the asymptotic bias of the 2SLS estimator, relative to that of the OLS estimator again only depends on k_z and μ. Furthermore, the distributions of the 2SLS t-ratio and Wald statistic only depend on μ, k_z and $\rho_{u\xi}$.

Summarizing, conventional first-order fixed-parameter asymptotics fail to give accurate approximations in case of weak instruments. Inspired by Bound, Jaeger and Baker (1995) and Staiger and Stock (1997) we use the concentration parameter to characterize relative bias and size distortions of Wald tests. One can proceed either with higher-order fixed-parameter asymptotics or consider weak instrument asymptotics. In the analysis below we have chosen for the former approach. In the panel AR(1) model weak instruments arise when $\alpha \to 1$ and/or $\frac{\sigma_\eta^2}{\sigma_v^2} \to \infty$. Kruiniger (2009) applies 'local to unity' asymptotics and shows that the Staiger and Stock (1997) set up does not always apply straigthforwardly to dynamic panel data models. More importantly, we find in our cross-sectional simulations below a weak instrument problem already for $\alpha = 0.4$ and $\frac{\sigma_\eta^2}{\sigma_v^2} = 4$, with the relative bias well approximated by (17). Expansion (15) also allows us to approximate the bias for less straightforward cases, like the cross-sectional system 2SLS estimator.

4 Cross section results for the AR(1) panel data model

Although the data are not generated as in the cross-section model (13), we can write the structural equation and the reduced form model for the AR(1) panel data model in first differences for the cross-section at time t as

$$\begin{aligned}\Delta y_{it} &= \alpha \Delta y_{i,t-1} + \Delta u_{it} \\ \Delta y_{i,t-1} &= y_i^{t-2\prime} \pi_{dt} + d_{i,t-1}.\end{aligned}$$

For the general expression of the expected value of the concentration parameter divided by n we get

$$E\left(\frac{1}{n}\mu_{dt}\right) = \frac{\pi_{dt}^{\prime} E\left(y_i^{t-2} y_i^{t-2\prime}\right) \pi_{dt}}{\sigma_{dt}^2}.$$

For the model in levels we have for the cross-section at time t

$$\begin{aligned}y_{it} &= \alpha y_{i,t-1} + \eta_i + v_{it} \\ y_{i,t-1} &= \Delta y_i^{t-1\prime} \pi_{lt} + l_{i,t-1}\end{aligned}$$

and the expected concentration parameter is given by

$$E\left(\frac{1}{n}\mu_{lt}\right) = \frac{\pi_{lt}^{\prime} E\left(\Delta y_i^{t-1} \Delta y_i^{t-1\prime}\right) \pi_{lt}}{\sigma_{lt}^2}.$$

In the Appendix we show that, under covariance stationarity of the initial observation,

$$E\left(\frac{1}{n}\mu_{dt}\right) = \frac{(1-\alpha)^2 \left(\sigma_v^2 + (t-3)\,\sigma_\eta^2\right)}{(1-\alpha^2)\,\sigma_v^2 + ((t-1)-(t-3)\,\alpha)\,(1+\alpha)\,\sigma_\eta^2}$$

and

$$E\left(\frac{1}{n}\mu_{lt}\right)=\frac{(t-2)(1-\alpha)^2\sigma_v^2}{(1-\alpha^2)\sigma_v^2+((t-1)-(t-3)\alpha)(1+\alpha)\sigma_\eta^2},$$

from which it follows that

$$\begin{aligned}\frac{E\left(\frac{1}{n}\mu_{dt}\right)}{E\left(\frac{1}{n}\mu_{lt}\right)} &= \frac{\left(\sigma_v^2+(t-3)\sigma_\eta^2\right)}{(t-2)\sigma_v^2}\\ &= \frac{1}{t-2}\left(1+(t-3)\frac{\sigma_\eta^2}{\sigma_v^2}\right).\end{aligned}$$

Therefore

$$E\left(\frac{1}{n}\mu_{dt}\right)=E\left(\frac{1}{n}\mu_{lt}\right)\quad \text{if } t=3,$$

and for $t>3$

$$\begin{aligned} E\left(\frac{1}{n}\mu_{dt}\right) &> E\left(\frac{1}{n}\mu_{lt}\right) \quad \text{if } \sigma_\eta^2>\sigma_v^2\\ E\left(\frac{1}{n}\mu_{dt}\right) &= E\left(\frac{1}{n}\mu_{lt}\right) \quad \text{if } \sigma_\eta^2=\sigma_v^2\\ E\left(\frac{1}{n}\mu_{dt}\right) &< E\left(\frac{1}{n}\mu_{lt}\right) \quad \text{if } \sigma_\eta^2<\sigma_v^2.\end{aligned}$$

Figure 1 graphs the values of $E\left(\frac{1}{n}\mu_{dt}\right)$ and $E\left(\frac{1}{n}\mu_{lt}\right)$ as a function of α for $t=6$ and various values of $\frac{\sigma_\eta^2}{\sigma_v^2}=\left\{\frac{1}{4},1,4\right\}$. The values of the concentration parameters decrease with increasing α. The concentration parameter for the LEV model is much more sensitive to the value of the variance ratio $\frac{\sigma_\eta^2}{\sigma_v^2}$ than the concentration parameter of the DIF model.

4.1 Discussion

The fact that the concentration parameters are the same in expectation for the IV estimators based on the DIF or LEV moment conditions for $t=3$ and for $t>3$ when $\sigma_\eta^2=\sigma_v^2$ seems contrary to the findings in Monte Carlo studies, see e.g. Blundell and Bond (1998) and Blundell, Bond and Windmeijer (2000) who use a covariance stationary design with $\sigma_\eta^2=\sigma_v^2=1$. In those simulation studies $\widehat{\alpha}_l$ outperforms $\widehat{\alpha}_d$ in terms of bias and rmse, especially when the series become more persistent, i.e. when α gets

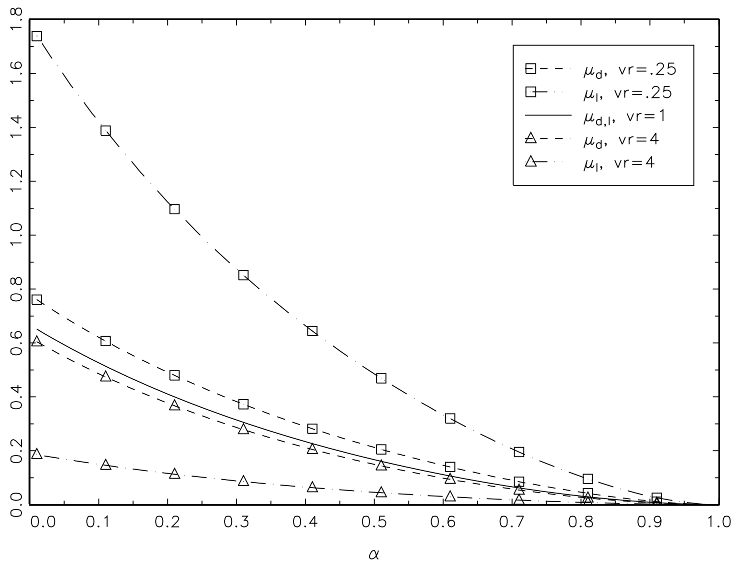

Figure 1: $E\left(\frac{1}{n}\mu\right)$ as a function of α, $t = 6$ and $vr = \frac{\sigma_\eta^2}{\sigma_v} = \left\{\frac{1}{4}, 1, 4\right\}$.

larger. The identification problem is apparent in the DIF model, where the reduced form parameters approach zero when α approaches 1. This is in sharp contrast to the reduced form parameters in the LEV model that approach $\frac{1}{2}$ when α approaches 1. This was the argument used by Blundell and Bond (1998) to assert the strength of the LEV moment conditions for the estimation of α for larger values of α.

There are two questions to be addressed. Firstly, why are the behaviours of the two estimators so different in terms of bias and rmse when they have the same expected concentration parameter? Secondly, how does the weak instrument problem in the LEV model manifest itself?

To answer the first question one has to realise that the structural models are different for DIF and LEV, with different endogeneity problems. Therefore, different biases arise for both OLS and 2SLS estimators in the two equations. For the DIF model

$$\Delta y_{it} = \alpha \Delta y_{i,t-1} + \Delta u_{it},$$

the OLS estimator for the cross-section at time t is given by

$$\widehat{\alpha}_{dOLS} = \alpha + \frac{\Delta y'_{t-1}\Delta u_t}{\Delta y'_{t-1}\Delta y_{t-1}},$$

and the limiting bias of the OLS estimator is, again assuming covariance stationarity,

$$\text{plim}\left(\widehat{\alpha}_{dOLS} - \alpha\right) = -\frac{1+\alpha}{2}.$$

For the LEV model

$$y_{it} = \alpha y_{i,t-1} + \eta_i + v_{it},$$

the OLS estimator is given by

$$\widehat{\alpha}_{lOLS} = \alpha + \frac{y'_{t-1}u_t}{y'_{t-1}y_{t-1}},$$

and the limiting bias of the OLS estimator is given by

$$\text{plim}\left(\widehat{\alpha}_{lOLS} - \alpha\right) = (1-\alpha)\frac{\frac{\sigma_\eta^2}{\sigma_v^2}}{\frac{\sigma_\eta^2}{\sigma_v^2} + \frac{1-\alpha}{1+\alpha}}$$

which reduces to $\text{plim}\left(\widehat{\alpha}_{lOLS} - \alpha\right) = \left(1-\alpha^2\right)/2$ when $\sigma_\eta^2 = \sigma_v^2$. The asymptotic absolute bias of $\widehat{\alpha}_{lOLS}$ is therefore (much) smaller than that of $\widehat{\alpha}_{dOLS}$ for high values of α.

Using (16) we can approximate the bias of the 2SLS estimator in the DIF model by

$$E\left(\widehat{\alpha}_d\right) - \alpha \approx \left(k_z - 2\right)\frac{\sigma_{\Delta u,d}}{\sigma_d^2}/E\left(\mu_d\right) = (t-4)\frac{\sigma_{\Delta u,d}}{\sigma_d^2}/E\left(\mu_d\right), \tag{18}$$

where we have suppressed the subscripts t for ease of exposition, and where

$$\sigma_{\Delta u,d} = E\left(\left(\Delta y_{i,t-1} - y_i^{t-2}\pi_d\right)\Delta u_{it}\right) = -\sigma_v^2.$$

Therefore

$$\begin{aligned} E\left(\widehat{\alpha}_d\right) - \alpha &\approx -(t-4)\,\sigma_v^2/\frac{n\sigma_v^2}{(1+\alpha)^2}\left(1-\alpha^2 - \frac{\sigma_\eta^2(1+\alpha)^2}{\sigma_v^2 + \sigma_\eta^2\left(t-3+\frac{1+\alpha}{1-\alpha}\right)}\right) \\ &= -(t-4)\frac{(1+\alpha)^2}{n\left(1-\alpha^2 - \frac{\sigma_\eta^2(1+\alpha)^2}{\sigma_v^2+\sigma_\eta^2\left(t-3+\frac{1+\alpha}{1-\alpha}\right)}\right)}, \end{aligned}$$

where we have used the expressions for σ_d^2 and $E(\mu_d)$ from the Appendix. Equivalently for the LEV model we get

$$E\left(\widehat{\alpha}_l\right) - \alpha \approx (t-4)\frac{\sigma_{u,l}}{\sigma_l^2}/E\left(\mu_l\right) \tag{19}$$

with

$$\sigma_{u,l} = E\left(\left(y_{i,t-1} - \Delta y_i^{t-1}\pi_l\right)u_{it}\right) = \frac{\sigma_\eta^2}{1-\alpha},$$

and therefore

$$\begin{aligned} E\left(\widehat{\alpha}_l\right) - \alpha &\approx \frac{(t-4)\,\sigma_\eta^2}{1-\alpha}/\frac{n\,(t-2)\,\sigma_v^2}{(1+\alpha)\left((t-1)-(t-3)\,\alpha\right)} \\ &= \frac{t-4}{t-2}\frac{\sigma_\eta^2}{\sigma_v^2}\frac{(1+\alpha)\left((t-1)-(t-3)\,\alpha\right)}{n\,(1-\alpha)}. \end{aligned}$$

Comparing these expressions is somewhat complicated but when $\sigma_\eta^2 = \sigma_v^2$ the absolute bias of the LEV 2SLS estimator will tend to be smaller than that of the DIF estimator. The main reason for this is that the absolute LEV OLS bias is smaller than the DIF OLS bias.

To answer the second question we now consider relative bias. Combining the results above on absolute OLS and 2SLS bias we get for the approximate relative absolute bias

$$\begin{aligned} \text{RelBias}_d &= \frac{\left|E\left(\widehat{\alpha}_d\right)-\alpha\right|}{\left|E\left(\widehat{\alpha}_{dOLS}-\alpha\right)\right|} \approx \frac{(t-4)\,E\left(\frac{1}{n}\mu_d\right)+1}{E\left(\mu_d\right)} \\ &= 2\,(t-4)\frac{(1+\alpha)}{n\left(1-\alpha^2-\frac{\sigma_\eta^2(1+\alpha)^2}{\sigma_v^2+\sigma_\eta^2\left(t-3+\frac{1+\alpha}{1-\alpha}\right)}\right)}, \end{aligned}$$

and

$$\begin{aligned} \text{RelBias}_l &= \frac{\left|E\left(\widehat{\alpha}_l\right)-\alpha\right|}{\left|E\left(\widehat{\alpha}_{lOLS}-\alpha\right)\right|} \approx \frac{(t-4)\,E\left(\frac{1}{n}\mu_l\right)+1}{E\left(\mu_l\right)} \\ &= \frac{t-4}{t-2}\frac{\left(\frac{\sigma_\eta^2}{\sigma_v^2}+\frac{1-\alpha}{1+\alpha}\right)(1+\alpha)\left((t-1)-(t-3)\,\alpha\right)}{n\,(1-\alpha)^2}. \end{aligned}$$

When $\sigma_\eta^2 = \sigma_v^2$ we have that $E\left(\frac{1}{n}\mu_d\right) = E\left(\frac{1}{n}\mu_l\right)$ and we expect therefore that the relative biases are the same for the DIF and LEV 2SLS estimators. Indeed this is the case and it amounts to

$$\text{RelBias}_d = \text{RelBias}_l \approx \frac{2\,(t-4)}{t-2}\frac{((t-1)-(t-3)\,\alpha)}{n\,(1-\alpha)^2}.$$

Finally, as mentioned in Section 3, the finite sample behaviour of the Wald test depends on the magnitude of the concentration parameter, the number of instruments and the correlation between the model errors. It is easily verified that $\rho^2_{\Delta u,d} = \rho^2_{u,l}$ when $\sigma^2_\eta = \sigma^2_v$ and therefore the size distortions of the Wald test will be the expected to be same for the DIF and LEV estimators in that case. When $\sigma^2_\eta < \sigma^2_v$ we have that both $E(\mu_d) < E(\mu_l)$ and that $\rho^2_{\Delta u,d} > \rho^2_{u,l}$, and therefore the Wald size distortion is expected to be smaller for the LEV estimator in that case. It is expected to be smaller for the DIF estimator when $\sigma^2_\eta > \sigma^2_v$ as then both $E(\mu_d) > E(\mu_l)$ and $\rho^2_{\Delta u,d} < \rho^2_{u,l}$.

4.2 System Estimator

For the cross-section at time t the SYS estimator combines the moment conditions of the DIF and LEV estimators. The OLS estimator in the SYS "model"

$$\begin{pmatrix} \Delta y_{it} \\ y_{it} \end{pmatrix} = \alpha \begin{pmatrix} \Delta y_{i,t-1} \\ y_{i,t-1} \end{pmatrix} + \begin{pmatrix} \Delta u_{it} \\ u_{it} \end{pmatrix} \tag{20}$$

is given by

$$\widehat{\alpha}_{sOLS} = \left(\Delta y'_{t-1}\Delta y_{t-1} + y'_{t-1}y_{t-1}\right)^{-1}\left(\Delta y'_{t-1}\Delta y_t + y'_{t-1}y_t\right)$$

and is clearly a weighted average of the DIF and LEV OLS estimators

$$\widehat{\alpha}_{sOLS} = \widetilde{\gamma}\widehat{\alpha}_{dOLS} + (1-\widetilde{\gamma})\,\widehat{\alpha}_{lOLS}$$

where

$$\widetilde{\gamma} = \frac{\Delta y'_{t-1}\Delta y_{t-1}}{\Delta y'_{t-1}\Delta y_{t-1} + y'_{t-1}y_{t-1}}$$

and

$$\text{plim}\,(\widetilde{\gamma}) = \frac{1-\alpha}{\frac{3}{2}-\alpha+\frac{1}{2}\frac{\sigma^2_\eta}{\sigma^2_v}\frac{1+\alpha}{1-\alpha}}.$$

The bias of the OLS estimator will therefore behave like the bias of the LEV OLS estimator when $\alpha \to 1$ and/or $\sigma_\eta^2/\sigma_v^2 \to \infty$, as $\widetilde{\gamma} \to 0$ in these cases. The asymptotic bias of $\widehat{\alpha}_{sOLS}$ is given by

$$\text{plim}\left(\widehat{\alpha}_{sOLS} - \alpha\right) = \frac{\left(1-\alpha^2\right)\left(\alpha - 1 + \frac{\sigma_\eta^2}{\sigma_v^2}\right)}{\left(3-2\alpha\right)\left(1-\alpha\right) + \frac{\sigma_\eta^2}{\sigma_v^2}\left(1+\alpha\right)}.$$

We can express the limiting bias of the SYS OLS estimator as

$$\text{plim}\left(\widehat{\alpha}_{sOLS}\right) - \alpha = \frac{\left(\sigma_{\Delta u,d} + \sigma_{ul}\right) / \left(\sigma_d^2 + \sigma_l^2\right)}{E\left(\frac{1}{n}\mu_s\right) + 1}$$

where

$$E\left(\frac{1}{n}\mu_s\right) = \phi E\left(\frac{1}{n}\mu_d\right) + \left(1-\phi\right) E\left(\frac{1}{n}\mu_l\right)$$

and

$$\phi = \frac{\sigma_d^2}{\left(\sigma_d^2 + \sigma_l^2\right)}.$$

When $\sigma_\eta^2 = \sigma_v^2$, we have then that $E\left(\frac{1}{n}\mu_s\right) = E\left(\frac{1}{n}\mu_d\right) = E\left(\frac{1}{n}\mu_l\right)$. As

$$\sigma_{\Delta u,d} + \sigma_{ul} = \frac{\sigma_\eta^2}{1-\alpha} - \sigma_v^2$$

we see that the absolute SYS OLS bias is then (substantially) smaller than the DIF and LEV OLS biases, and equal to 0 when $\alpha = 0$.

Figure 2 shows the asymptotic biases of the DIF, LEV and SYS OLS estimators as a function of α for different values of $\sigma_\eta^2/\sigma_v^2 = \left\{\frac{1}{4}, 1, 4\right\}$. It is clear from this picture that the LEV and SYS OLS biases are much smaller than the DIF OLS bias for higher values of α.

The SYS 2SLS estimator for cross section t is also a weighted average of the DIF and LEV cross sectional 2SLS estimators

$$\widehat{\alpha}_s = \widetilde{\delta}\widehat{\alpha}_d + \left(1 - \widetilde{\delta}\right)\widehat{\alpha}_l$$

where

$$\widetilde{\delta} = \frac{\widehat{\pi}_d' Z_d' Z_d \widehat{\pi}_d}{\widehat{\pi}_d' Z_d' Z_d \widehat{\pi}_d + \widehat{\pi}_l' Z_l' Z_l \widehat{\pi}_l},$$

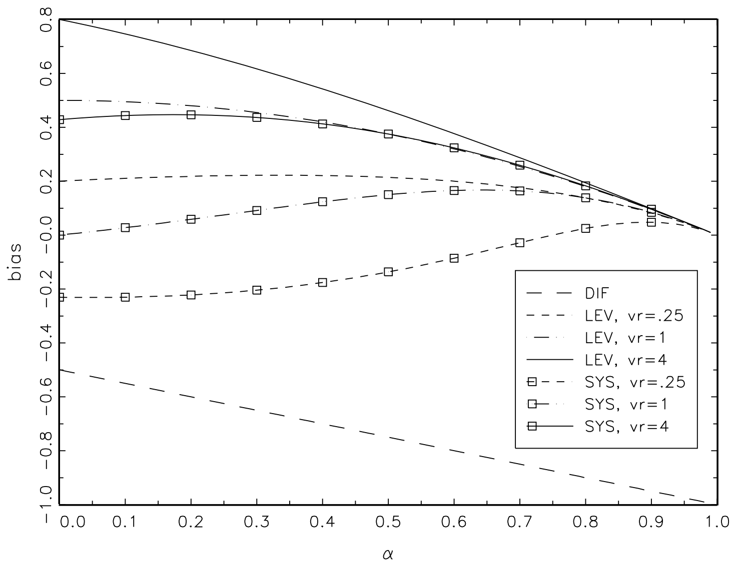

Figure 2: Asymptotic biases of OLS estimators, $vr = \sigma_\eta^2/\sigma_v^2 = \left\{\frac{1}{4}, 1, 4\right\}$.

see also Blundell, Bond and Windmeijer (2000), with

$$\operatorname{plim}\left(\widetilde{\delta}\right) = \frac{E\left(\frac{1}{n}\mu_d\right)}{E\left(\frac{1}{n}\mu_d\right) + \frac{\sigma_l^2}{\sigma_d^2} E\left(\frac{1}{n}\mu_l\right)}$$

and again $\widetilde{\delta} \to 0$ if $\alpha \to 1$ and/or $\sigma_\eta^2/\sigma_v^2 \to \infty$. Clearly, the absolute bias of the SYS 2SLS estimator will be smaller than the maximum of the absolute biases of the DIF and LEV 2SLS estimators.

Combining the results of the OLS biases, values of the concentration parameters in the DIF and LEV models and relative weights on the DIF and LEV moment conditions in the SYS 2SLS estimator, we expect the absolute bias of the SYS estimator to be small for large values of α, but that this bias is an increasing function of $\frac{\sigma_\eta^2}{\sigma_v^2}$. This happens because the bias of the LEV OLS estimator is an increasing function of $\frac{\sigma_\eta^2}{\sigma_v^2}$, the LEV concentration parameter a decreasing function of $\frac{\sigma_\eta^2}{\sigma_v^2}$, and the weight $\left(1 - \widetilde{\delta}\right)$ an increasing function in $\frac{\sigma_\eta^2}{\sigma_v^2}$, implying that more weight will be given to the LEV moment conditions.

The definition of μ_s above suggest a concentration parameter equivalent for the SYS model given by

$$\mu_s = \frac{\pi_d' Z_d' Z_d \pi_d + \pi_l' Z_l' Z_l \pi_l}{\sigma_d^2 + \sigma_l^2}.$$

However, in this case the value of μ_s does not directly convey the magnitude of the bias of the 2SLS estimator, relative to the bias of the OLS estimator. This is due to the additional covariance terms of the reduced form errors d and l. As in (15), consider the approximation

$$\begin{aligned} E\left(n^{1/2}\left(\widehat{\alpha}_s - \alpha\right)\right) &\approx E\left(\frac{\pi_d' z_{d,\Delta u} + \pi_l' z_{l,u}}{\pi_d' Q_d \pi_d + \pi_l' Q_l \pi_l}\right) \\ &\quad + n^{-1/2} E\left(\frac{z_{d,d}' Q_d z_{d,\Delta u} + z_{l,l}' Q_l z_{l,u}}{\pi_d' Q_d \pi_d + \pi_l' Q_l \pi_l}\right) \\ &\quad - 2n^{-1/2} E\left(\frac{\left(\pi_d' z_{d,d} + \pi_l' z_{l,l}\right)\left(\pi_d' z_{d,\Delta u} + \pi_l' z_{l,u}\right)}{\left(\pi_d' Q_d \pi_d + \pi_l' Q_l \pi_l\right)^2}\right), \end{aligned}$$

where $z_{a,b} = \frac{1}{\sqrt{n}} Z_a' b$. We then get the approximate bias expression for the SYS 2SLS estimator:

$$\begin{aligned} E\left(\widehat{\alpha}_s\right) - \alpha &\approx \frac{1}{n} \frac{(t-2)\left(\sigma_{\Delta u,d} + \sigma_{ul}\right)}{\pi_d' Q_d \pi_d + \pi_l' Q_l \pi_l} \\ &\quad - \frac{2}{n} \frac{\sigma_\eta^2 / (1-\alpha)\, \pi_l' Q_l \pi_l - \sigma_v^2 \pi_d' Q_d \pi_d}{\left(\pi_d' Q_d \pi_d + \pi_l' Q_l \pi_l\right)^2} \\ &\quad - \frac{2}{n} E\left(\frac{\left(\pi_d' z_{d,d}\right)\left(\pi_l' z_{l,u}\right) + \left(\pi_l' z_{l,l}\right)\left(\pi_d' z_{d,\Delta u}\right)}{\left(\pi_d' Q_d \pi_d + \pi_l' Q_l \pi_l\right)^2}\right) \end{aligned} \qquad (21)$$

We calculate this approximate bias expression and the associated relative bias for the Monte Carlo simulation example in the next section, where it is shown that the relative bias of the SYS estimator is smaller than that of the LEV or DIF estimator when $\sigma_\eta^2 = \sigma_v^2$ even though in that case $E(\mu_d) = E(\mu_l) = E(\mu_s)$.

Clearly, the SYS 2SLS estimator is not efficient as there is heteroskedasticity and correlation between the errors in model (20). We will focus on the 2SLS estimator here in the cross-section analysis and consider the efficient 2-step GMM estimator below when considering the full panel data analysis.

4.3 Some Monte Carlo Results

To investigate the finite sample behaviour of the estimators and Wald test statistics we conduct the following Monte Carlo experiment. We compute the OLS and 2SLS estimators for LEV, DIF and SYS for the cross section $t = 6$ for the model specification

$$\begin{aligned} y_{i1} &= \frac{\eta_i}{1-\alpha} + \varepsilon_i; \\ y_{it} &= \alpha y_{i,t-1} + \eta_i + v_{it}; \end{aligned}$$

$$\varepsilon_i \sim n\left(0, \frac{\sigma_v^2}{1-\alpha^2}\right);\ \eta_i \sim N\left(0, \sigma_\eta^2\right);\ v_{it} \sim N\left(0, \sigma_v^2\right),$$

for sample size $n = 200$; $\sigma_v^2 = 1$, and different values of $\alpha = \{0.4, 0.8\}$ and $\sigma_\eta^2 = \left\{\frac{1}{4}, 1, 4\right\}$. Note that in this design results depend only on the relative value $vr = \sigma_\eta^2/\sigma_v^2$, not the total variance $\sigma_\eta^2 + \sigma_v^2$. There are 4 instruments for the DIF and LEV 2SLS estimators, whereas the SYS 2SLS estimator is in this cross-sectional case based on the 8 combined moment conditions. Tables 1 and 2 present the estimation results for $10,000$ Monte Carlo replications.

The results in Tables 1 and 2 confirm the findings and conjectures stated in the previous sections. The DIF OLS (absolute) bias is larger than the LEV OLS bias in all cases, especially when the series are more persistent when $\alpha = 0.8$. The relative biases of the DIF and LEV 2SLS estimators are, however, the same when $vr = \sigma_\eta^2/\sigma_v^2 = 1$. These relative biases are equal to 0.052 and 0.057 respectively when $\alpha = 0.4$, in which case the expected concentration parameters are equal to 46.75. The relative biases are larger, 0.310 and 0.312 respectively when $\alpha = 0.8$. For this case the expected concentration parameters are much smaller and equal to 6.35, which corresponds to a first-stage F-statistic of $6.35/4 = 1.58$.

The relative bias of the DIF 2SLS estimator does not vary much with the different values of vr when $\alpha = 0.4$, whereas that of the LEV 2SLS estimator does. It is only 0.029 when $vr = 1/4$, but increases to 0.169 when $vr = 4$. These are exactly in line with the larger variation in the values of the expected concentration parameter for the LEV model. They are 132.7 when $vr = 1/4$ and 13.0 when $vr = 4$, compared to 58.1 and 42.3 respectively for the DIF model. The absolute bias of the DIF 2SLS estimator is smaller than that of the LEV 2SLS one when $vr = 4$, but larger in the other cases.

When $\alpha = 0.8$, there is a similar pattern to the results of the relative biases. For the LEV 2SLS model it now decreases to 0.11 when $vr = 1/4$, with the expected concentration parameter equal to 20.9. It increases to 0.68 when $vr = 4$ and the expected concentration parameter is only 1.68. As explained before, we see that the weak instrument problem for the LEV moment conditions, given α, becomes more severe with increasing vr. As both the OLS bias and the relative bias increase with increasing vr, so does the absolute bias of the 2SLS estimator. When $\alpha = 0.8$, the absolute bias of the LEV 2SLS estimator ranges from 0.015 when $vr = 1/4$ to 0.132 when $vr = 4$.

The SYS 2SLS estimator has a slightly smaller relative bias than the DIF and LEV ones when $vr = 1$. It is 0.03 when $\alpha = 0.4$ and 0.24 when $\alpha = 0.8$. Unlike the results for the LEV 2SLS estimator, the relative bias actually increases when $vr = 1/4$, although the absolute bias is quite small, especially when $\alpha = 0.8$. The relative bias is quite large in that case because the bias of the SYS OLS estimator is very small. When $vr = 4$ the relative and absolute biases of the SYS 2SLS estimator are similar to that of the LEV 2SLS estimator, albeit slightly smaller.

Table 2 further shows that the higher order bias and relative 2SLS bias approximations calculated from (16) and (17) for DIF and LEV and from (21) for SYS are very accurate. The exception is when the concentration parameter is very small for LEV when $\alpha = 0.8$ and $vr = 4$. Then the bias approximations indicate too high a bias for LEV and SYS.

Figures 3 and 4 display p-value plots for the Wald test for testing $H_0 : \alpha = \alpha_0$ with α_0 the true parameter value. When $vr = 1$, the size properties of the Wald tests based on the DIF and LEV 2SLS estimates are virtually identical, which is as expected as the

Table 1: Cross Section Estimation Results for $n = 200$ and $t = 6$

			DIF		LEV		SYS	
α	vr		Coeff	*StDev*	Coeff	*StDev*	Coeff	*StDev*
0.4	1/4	OLS	-0.300	*0.067*	0.621	*0.056*	0.224	*0.057*
		2SLS	0.370	*0.173*	0.406	*0.092*	0.389	*0.081*
		$E(\mu)$	58.06		132.7			
	1	OLS	-0.301	*0.067*	0.820	*0.041*	0.523	*0.049*
		2SLS	0.364	*0.189*	0.424	*0.113*	0.404	*0.095*
		$E(\mu)$	46.75		46.75			
	4	OLS	-0.301	*0.067*	0.942	*0.024*	0.812	*0.029*
		2SLS	0.360	*0.197*	0.492	*0.157*	0.462	*0.122*
		$E(\mu)$	42.31		13.02			
0.8	1/4	OLS	-0.100	*0.070*	0.938	*0.025*	0.824	*0.028*
		2SLS	0.597	*0.404*	0.815	*0.084*	0.793	*0.083*
		$E(\mu)$	9.15		20.92			
	1	OLS	-0.100	*0.070*	0.980	*0.014*	0.938	*0.015*
		2SLS	0.521	*0.464*	0.856	*0.092*	0.834	*0.090*
		$E(\mu)$	6.35		6.35			
	4	OLS	-0.100	*0.070*	0.995	*0.007*	0.983	*0.007*
		2SLS	0.484	*0.485*	0.932	*0.085*	0.917	*0.079*
		$E(\mu)$	5.45		1.68			

Notes: Means and standard deviations of 10,000 estimates. $vr = \sigma_\eta^2/\sigma_v^2$.

Table 2: Bias approximations $n = 200$ and $t = 6$

α	vr	DIF Bias	DIF RelBias	LEV Bias	LEV RelBias	SYS Bias	SYS RelBias
0.4	1/4	-0.030	0.043	0.006	0.029	-0.011	0.063
		-0.031	*0.044*	*0.006*	*0.025*	*-0.012*	*0.068*
	1	-0.036	0.052	0.024	0.057	0.004	0.031
		-0.037	*0.053*	*0.022*	*0.053*	*0.003*	*0.021*
	4	-0.039	0.057	0.092	0.169	0.062	0.151
		-0.040	*0.057*	*0.089*	*0.164*	*0.065*	*0.157*
0.8	1/4	-0.203	0.225	0.015	0.109	-0.007	0.314
		-0.206	*0.229*	*0.015*	*0.106*	*-0.010*	*0.403*
	1	-0.279	0.310	0.056	0.312	0.034	0.243
		-0.293	*0.325*	*0.059*	*0.325*	*0.033*	*0.241*
	4	-0.316	0.351	0.132	0.681	0.117	0.640
		-0.339	*0.377*	*0.234*	*1.203*	*0.208*	*1.140*

Notes: Mean bias and relative bias from 10,000 estimates. RelBias= $\left|\bar{\hat{\alpha}}_{2SLS} - \alpha\right| / \left|\bar{\hat{\alpha}}_{OLS} - \alpha\right|$. Higher order bias approximations in *italics*. $vr = \sigma_\eta^2/\sigma_v^2$.

concentration parameters are equal in expectation as are the correlation coefficients of the model errors. It is also clear that when $\alpha = 0.8$, the size properties of the Wald tests are very poor, with a large overrejection of the null reflecting the low value of the concentration parameters. The size properties of the Wald test based on the SYS 2SLS estimation results are better than those based on the DIF and LEV 2SLS results, but again very poor when $\alpha = 0.8$. When $vr = 1/4$ the size properties of the Wald tests based on the LEV and SYS 2SLS estimation results are quite good, even when $\alpha = 0.8$, whereas they are very poor when $vr = 4$. The Wald test results based on the DIF 2SLS estimates are not very sensitive to the value of vr. These results are again in line with expectation given the results of the previous section.

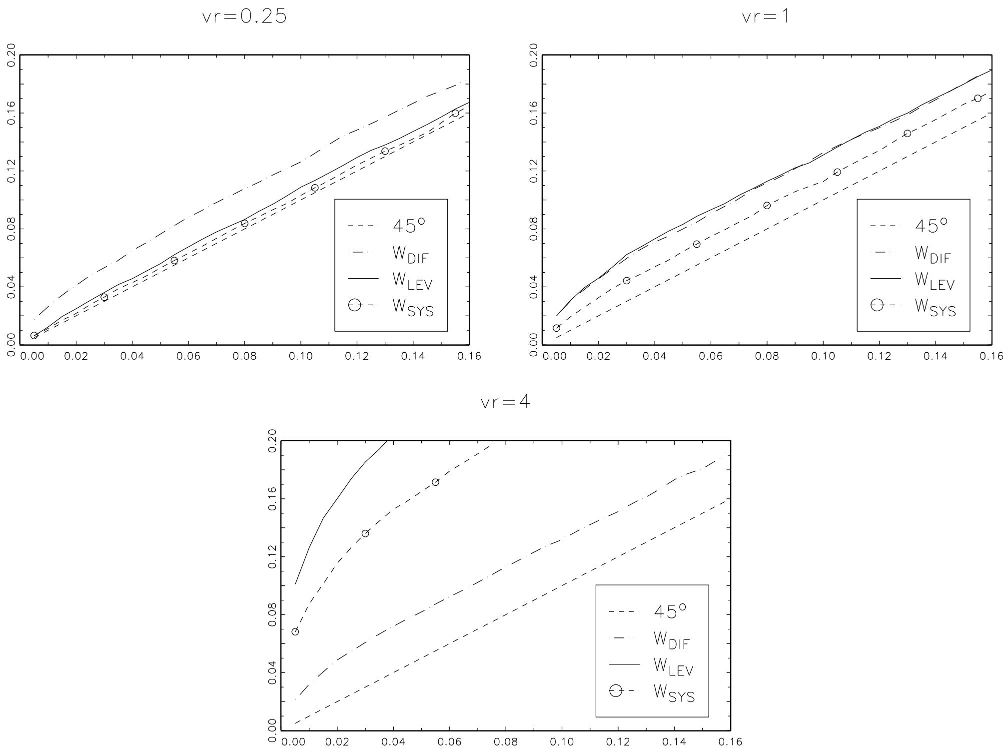

Figure 3: P-value plots, Wald test, $H_0 : \alpha = 0.4$; $vr = \sigma_\eta^2/\sigma_v^2$.

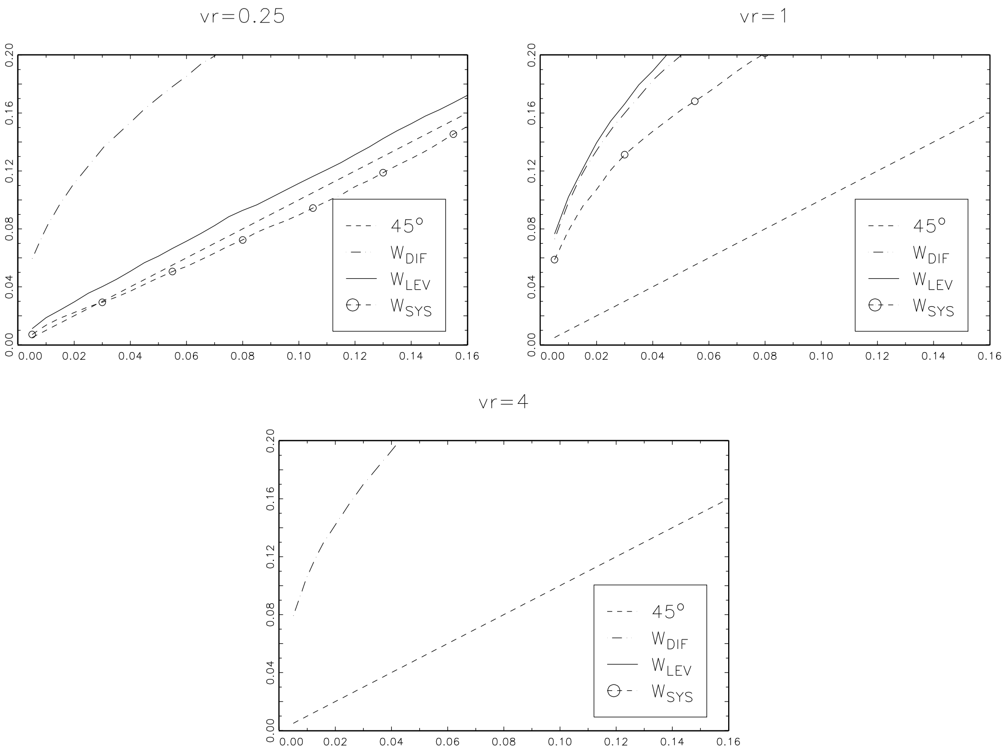

Figure 4: P-value plots, Wald test, $H_0 : \alpha = 0.8$; $vr = \sigma_\eta^2/\sigma_v^2$.

4.4 Mean Stationarity Only

In all the derivations so far we assumed covariance stationarity of the initial condition. When we assume mean stationarity only, i.e.

$$y_{i1} = \frac{\eta_i}{1-\alpha} + \varepsilon_i$$

with $E\left(\varepsilon_i^2\right) = \sigma_\varepsilon^2$, we show in the Appendix that for $t = 3$

$$\begin{aligned} E\left(\frac{1}{n}\mu_{l3}\right) &> E\left(\frac{1}{n}\mu_{d3}\right) \text{ if } \sigma_\varepsilon^2 < \frac{\sigma_v^2}{1-\alpha^2} \\ E\left(\frac{1}{n}\mu_{l3}\right) &< E\left(\frac{1}{n}\mu_{d3}\right) \text{ if } \sigma_\varepsilon^2 > \frac{\sigma_v^2}{1-\alpha^2}, \end{aligned}$$

so that, when $t = 3$, the expected concentration parameter for the LEV model is larger than that of the DIF model when the variance of the initial condition is smaller than the covariance stationary level and vice versa.

5 Panel Data Analysis

The concept of the concentration parameter and its relationship to relative bias and size distortion of the Wald test does not readily extend itself to general GMM estimation, see e.g. Stock and Wright (2000) and Han and Phillips (2006). Estimation of the panel AR(1) model by 2SLS, using all available time periods and the full set of sequential moment conditions for the DIF and SYS models (6) and (12) will result in a weighted average of the period specific 2SLS estimates. Weighting by the efficient weight matrix will lead to different results, but we expect the weak instrument issues as documented in the previous section for the DIF and LEV cross-sectional estimates to carry over to the linear GMM estimation. This is indeed confirmed by our Monte Carlo results presented here.

Tables 3 and 4 presents Monte Carlo estimation results for the AR(1) model with normally distributed η_i and v_i, with $n = 200$, $T = 6$, $\alpha = 0.8$ and $vr = (0.25, 1, 4)$. We

Table 3: Panel Data Estimation Results, $n = 200$, $T = 6$, $\alpha = 0.8$

	DIF		LEV		SYS	
	Coeff	*StDev*	Coeff	*StDev*	Coeff	*StDev*
$vr = 1/4$						
OLS	-0.100	*0.033*	0.938	*0.011*	0.824	*0.018*
2SLS	0.581	*0.162*	0.812	*0.056*	0.779	*0.074*
1-step	0.734	*0.131*			0.798	*0.067*
2-step	0.734	*0.140*	0.812	*0.060*	0.797	*0.060*
$vr = 1$						
OLS	-0.100	*0.033*	0.980	*0.006*	0.938	*0.009*
2SLS	0.469	*0.212*	0.850	*0.068*	0.813	*0.079*
1-step	0.672	*0.181*			0.830	*0.073*
2-step	0.664	*0.201*	0.844	*0.042*	0.818	*0.068*
$vr = 4$						
OLS	-0.100	*0.033*	0.995	*0.003*	0.983	*0.004*
2SLS	0.401	*0.240*	0.924	*0.069*	0.889	*0.075*
1-step	0.618	*0.213*			0.900	*0.070*
2-step	0.601	*0.241*	0.913	*0.079*	0.884	*0.079*

Note: Means and standard deviations of 10,000 estimates.

Table 4: Bias and Relative Bias, $\alpha = 0.8$, $T = 6$

	DIF		LEV		SYS	
vr	Bias	RelBias	Bias	RelBias	Bias	RelBias
1/4	-0.219	0.244	0.012	0.086	-0.021	0.887
1	-0.331	0.367	0.050	0.279	0.013	0.093
4	-0.399	0.443	0.124	0.637	0.089	0.488

Notes: Mean and relative bias from 10,000 estimates. RelBias= $\left|\bar{\hat{\alpha}}_{2SLS} - \alpha\right| / \left|\bar{\hat{\alpha}}_{OLS} - \alpha\right|$.

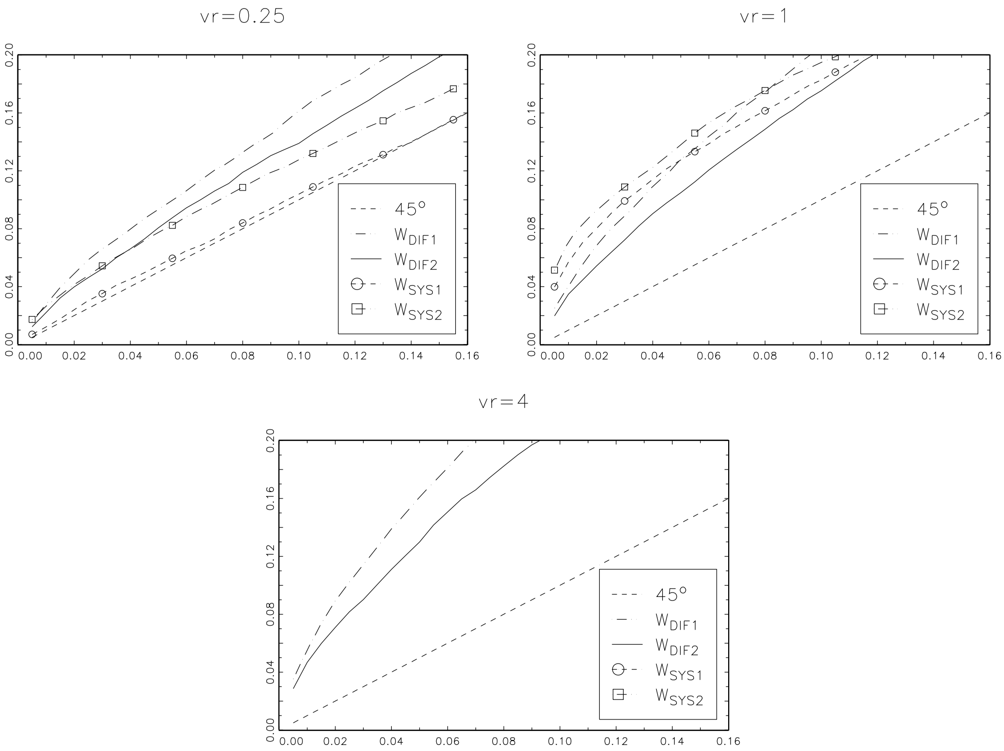

Figure 5: P-value plots, Wald test, $H_0 : \alpha = 0.8$, $T = 6$; $vr = \sigma_\eta^2/\sigma_v^2$.

present 2SLS and 1-step and 2-step GMM estimation results. We use for the initial weight matrix for the 1-step GMM DIF estimator $W_n = \sum_{i=1}^{n} Z'_{di} A Z_{di}$ where A is a $(T-2)$ square matrix that has 2s on the main diagonal, -1s on the first subdiagonals, and zeros elsewhere. This is the efficient weight matrix for the DIF moment conditions when the v_{it} are homoskedastic and not serially correlated, as is the case here. For the 1-step GMM SYS estimator we use the commonly used initial weight matrix $W_n = \sum_{i=1}^{n} Z'_{si} H Z_{si}$ where H is a $2(T-2)$ square matrix

$$H = \begin{bmatrix} A & 0 \\ 0 & I_{T-2} \end{bmatrix},$$

where I_{T-2} is the identity matrix of order $T-2$.

The pattern of results for the 2SLS estimates is quite similar to that found for the $t = 6$ cross-section as reported in Table 1. The DIF 2SLS estimator displays somewhat larger relative biases, whereas the LEV 2SLS estimator has smaller relative biases than in the cross-section. SYS has smaller relative and absolute biases at $vr = 1$ and $vr = 4$, but the direction of the biases remain the same.

Use of the efficient initial weight matrix reduces the bias of the 1-step GMM DIF estimator significantly. This is due to the fact that the comparison bias is now no longer the OLS bias in the first differenced model, but the bias of the within groups estimator, which is smaller. There is no clear pattern to the bias of the SYS one- and two-step GMM estimators in comparison to the 2SLS estimator.

Figure 5 displays the p-value plots of the Wald tests for testing $H_0 : \alpha = 0.8$ based on the DIF and SYS GMM estimation results, where the Wald tests based on the 2-step GMM results use the Windmeijer (2005) corrected variance estimates. The pattern of size properties is very similar to that for the cross-section analysis. The Wald test based on the SYS GMM estimation results has better size properties than that based on the DIF GMM estimation results when $vr = 0.25$, especially for the 1-step SYS GMM estimator. The size behaviours are very similar when $vr = 1$, but the SYS Wald tests size properties are much worse than that of the DIF Wald tests when $vr = 4$.

As for the cross-sectional SYS estimator, we can start with the bias of the panel DIF OLS estimator in order to obtain a suggestion for a concentration parameter.

$$\operatorname{plim}\left(\widehat{\alpha}_{dOLS}\right) - \alpha = \frac{-\left(T-2\right)\sigma_v^2}{\sum_{t=3}^{T}\pi_{dt}' Q_{dt}\pi_{dt} + \sum_{t=3}^{T}\sigma_{dt}^2},$$

suggesting a concentration parameter defined as

$$\mu_d = \frac{\sum_{t=3}^{T}\pi_{dt}' Z_{dt}' Z_{dt}\pi_{dt}}{\sum_{t=3}^{T}\sigma_{dt}^2}.$$

For the 2SLS bias we get

$$\begin{aligned} E\left(n^{1/2}\left(\widehat{\alpha}_d - \alpha\right)\right) &\approx E\left(\frac{\pi_d' z_{d,\Delta u}}{\pi_d' Q_d \pi_d}\right) \\ &\quad + n^{-1/2}\left(E\left(\frac{z_{d,d}' Q_d z_{d,\Delta u}}{\pi_d' Q_d \pi_d}\right) - 2E\left(\frac{\left(\pi_d' z_{d,d}\right)\left(\pi_d' z_{d,\Delta u}\right)}{\left(\pi_d' Q_d \pi_d\right)^2}\right)\right), \end{aligned}$$

$$\begin{aligned} E\left(\widehat{\alpha}_d\right) - \alpha &\approx -\frac{1}{n}\frac{\left(\left(T-1\right)\left(T-2\right)/2-2\right)\sigma_v^2}{\pi_d' Q_d \pi_d} \\ &\quad - \frac{2}{n}E\left(\frac{\sum_{t=3}^{T}\sum_{j\neq t}\left(\pi_{dt}' z_{d,dt}\right)\left(\pi_{dj}' z_{d,\Delta uj}\right)}{\left(\pi_d' Q_d \pi_d\right)^2}\right) \end{aligned}$$

As before for the SYS cross-sectional 2SLS estimator, the concentration parameter μ_{pd} does not convey all the information concerning the relative bias of the 2SLS estimator, due to the additional covariance terms in the expansion. Equivalent results can be obtained for the panel LEV and panel SYS 2SLS estimators. For the efficient one-step panel DIF GMM estimator, similar expansions can be derived, but now for the model where the individual data in the model is premultiplied by $A^{-1/2}$, but the instruments by $A^{1/2}$.

5.1 Bias Approximations for Panel 2SLS Estimators

Although the concept of concentration parameter does not automatically extend to panels it is possible to analyse absolute and relative bias of panel estimators of α. We now consider panel IV estimators, i.e. exploiting the identity weight matrix in the definitions of $\hat{\alpha}_d$ and $\hat{\alpha}_l$. Hence, the W_N matrix is of the simple form $Z'Z$. We analyze the DIF and LEV panel IV estimators using results from Alvarez and Arellano (2003) and Hayakawa (2008) respectively. In those studies probability limits of the DIF and LEV panel IV estimators have been derived assuming both T and N growing large with $T/N \to c$, $0 \leq c < \infty$. Regarding the panel DIF 2SLS estimator from Theorem 4 of Alvarez and Arellano (2003) we have

$$\text{plim}\,(\hat{\alpha}_d - \alpha) = -\frac{1+\alpha}{2}\left(\frac{c}{2-(1+\alpha)(2-c)/2}\right),$$

while for the panel LEV 2SLS estimator using Theorem 3 of Hayakawa (2007) we have

$$\text{plim}\,(\hat{\alpha}_l - \alpha) = \frac{\frac{c}{2}\frac{\sigma_\eta^2}{\sigma_v^2}\left(\frac{1}{1-\alpha}\right)}{\frac{c}{2}\frac{\sigma_\eta^2}{\sigma_v^2}\left(\frac{1}{1-\alpha}\right)^2 + \frac{1}{1-\alpha^2}}.$$

Hence, for both T and N large panel IV estimators are inconsistent. Comparing these asymptotic 2SLS biases with the limiting biases of OLS (see Section 4.1 for analytical expressions) we find that for $\frac{\sigma_\eta^2}{\sigma_v^2} = 1$ relative bias for DIF and LEV is equal and amounts to

$$\frac{c}{\frac{c}{2}(1+\alpha)+1-\alpha}.$$

Furthermore, relative bias for LEV is larger then DIF when $\frac{\sigma_\eta^2}{\sigma_v^2} > 1$ and vice versa. Hence, these results for panel IV estimators mimic the cross-sectional results on relative bias as discussed in Section 4.

Panel 2SLS estimators can be expressed as a weighted average of period specific 2SLS estimators. This suggests that cross-section based concentration parameters as derived in the previous section are also informative about absolute and relative 2SLS bias when exploiting the whole panel. This conjecture is correct as we will now show. The above results of Alvarez and Arellano (2003) and Hayakawa (2008) can be interpreted as the 2SLS inconsistency under many instrument asymptotics. Hence, the bias of panel 2SLS estimators when the number of instruments is reasonably large can be approximated by

$$E(\hat{\alpha}_d - \alpha) \approx \frac{E\left(\Delta y_{-1}' Z_d (Z_d' Z_d)^{-1} Z_d' \Delta u\right)}{E\left(\Delta y_{-1}' Z_d (Z_d' Z_d)^{-1} Z_d' \Delta y_{-1}\right)},$$

$$E(\hat{\alpha}_l - \alpha) \approx \frac{E\left(y_{-1}' Z_l (Z_l' Z_l)^{-1} Z_l' u\right)}{E\left(y_{-1}' Z_l (Z_l' Z_l)^{-1} Z_l' y_{-1}\right)}.$$

The above expressions are basically an evaluation of the expected value of the leading term (inconsistency) in an asymptotic expansion of the estimation error under many instruments. In the Appendix we show that

$$E(\hat{\alpha}_d - \alpha) \approx \frac{0.5(T-1)(T-2)\sigma_{\Delta u,d}}{\sum_{t=3}^{T} \sigma_{dt}^2 \left(E(\mu_{dt}) + (t-2)\right)},$$

$$E(\hat{\alpha}_l - \alpha) \approx \frac{0.5(T-1)(T-2)\sigma_{u,l}}{\sum_{t=3}^{T} \sigma_{lt}^2 \left(E(\mu_{lt}) + (t-2)\right)}.$$

Indeed cross-section specific concentration parameters appear in these bias approximations. Although analytically no tractable expression results it is interesting that regarding relative bias numerically the same pattern as in the pure cross-section case emerges. In other words, relative bias for panel DIF is larger than for panel LEV when $\sigma_\eta^2 < \sigma_v^2$ and vice versa. And when the variance ratio σ_η^2/σ_v^2 is equal to 1 we have that the relative biases for the estimators are equal.

Regarding the panel SYS 2SLS estimator we can proceed in a similar way and evaluate

$$E(\hat{\alpha}_s - \alpha) \approx \frac{E\left(q_{-1}' Z_s (Z_s' Z_s)^{-1} Z_s' p\right)}{E\left(q_{-1}' Z_s (Z_s' Z_s)^{-1} Z_s' q_{-1}\right)}.$$

In the Appendix we show that

$$E(\hat{\alpha}_s - \alpha) \approx \frac{0.5(T-1)(T-2)\sigma_{\Delta u,d} + (T-2)\sigma_{u,l}}{\sum_{t=3}^{T} \sigma_{dt}^2 (E(\mu_{dt}) + (t-2)) + \sum_{t=3}^{T} \sigma_{lt}^2 (E(\mu_{lt}) + 1)}.$$

We expect the bias approximations of the panel IV estimators to work well when at least T is moderately large compared with N. Table 5 presents estimation results for the panel data Monte Carlo exercise when $T = 15$. Table 6 further presents the bias

Table 5: Panel Data Estimation Results, $n = 200$, $T = 15$, $\alpha = 0.8$

	DIF		LEV		SYS	
	Coeff	*StDev*	Coeff	*StDev*	Coeff	*StDev*
$vr = 1/4$						
OLS	-0.100	*0.019*	0.938	*0.007*	0.824	*0.014*
2SLS	0.426	*0.069*	0.828	*0.024*	0.730	*0.041*
1-step	0.767	*0.034*			0.793	*0.029*
2-step	0.766	*0.039*	0.822	*0.027*	0.796	*0.027*
$vr = 1$						
OLS	-0.100	*0.019*	0.980	*0.003*	0.938	*0.006*
2SLS	0.374	*0.075*	0.880	*0.027*	0.776	*0.043*
1-step	0.757	*0.040*			0.819	*0.031*
2-step	0.754	*0.046*	0.866	*0.032*	0.816	*0.030*
$vr = 4$						
OLS	-0.100	*0.019*	0.995	*0.001*	0.983	*0.002*
2SLS	0.355	*0.078*	0.946	*0.023*	0.868	*0.039*
1-step	0.751	*0.042*			0.882	*0.031*
2-step	0.748	*0.048*	0.935	*0.031*	0.877	*0.033*

Notes: Means and standard deviations of 10,000 estimates; $vr = \sigma_\eta^2/\sigma_v^2$.

approximations. As expected, we now find that the relative biases of the DIF and LEV estimators are virtually identical for $T = 15$. We also include those for $T = 6$. These results corroborate our large T theoretical findings, with reasonable approximations even when $T = 6$, especially for DIF.

6 Conclusions

We have shown that the concentration parameters in the reduced forms of the DIF and LEV cross-sectional models are the same in expectation when the variances of the unobserved heterogeneity (σ_η^2) and idiosyncratic errors (σ_v^2) are the same in the covariance stationary AR(1) model. The LEV concentration parameter is smaller than the DIF one if $\sigma_\eta^2 > \sigma_v^2$ and it is larger if $\sigma_\eta^2 < \sigma_v^2$. Therefore, the well-understood weak instrument problem in the DIF model also applies to the LEV model, especially when $\sigma_\eta^2 \geq \sigma_v^2$,

Table 6: Panel Bias Approximations, $\alpha = 0.8$

		DIF		LEV		SYS	
T	vr	Bias	RelBias	Bias	RelBias	Bias	RelBias
6	1/4	-0.219	0.244	0.012	0.086	-0.021	0.887
		-0.227	*0.252*	*0.023*	*0.164*	*-0.017*	*0.673*
	1	-0.331	0.367	0.050	0.279	0.013	0.093
		-0.339	*0.377*	*0.068*	*0.377*	*0.028*	*0.203*
	4	-0.399	0.443	0.124	0.637	0.089	0.488
		-0.407	*0.453*	*0.134*	*0.691*	*0.107*	*0.583*
15	1/4	-0.374	0.416	0.028	0.200	-0.070	2.960
		-0.376	*0.418*	*0.031*	*0.227*	*-0.069*	*2.813*
	1	-0.426	0.473	0.080	0.445	-0.024	0.174
		-0.428	*0.475*	*0.086*	*0.475*	*-0.020*	*0.145*
	4	-0.445	0.495	0.146	0.752	0.068	0.370
		-0.447	*0.497*	*0.150*	*0.770*	*0.075*	*0.409*

Notes: Mean and relative bias from 10,000 estimates. RelBias= $\left|\overline{\widehat{\alpha}}_{2SLS} - \alpha\right| / \left|\overline{\widehat{\alpha}}_{OLS} - \alpha\right|$. Bias approximations in *italics*. $vr = \sigma_\eta^2/\sigma_v^2$.

with both concentration parameters decreasing in value with increasing persistence of the data series. The weak instrument problem does manifest itself in the magnitude of the bias of 2SLS relative to that of OLS, which we show are equal for DIF and LEV when $\sigma_\eta^2 = \sigma_v^2$. The LEV 2SLS estimator has a smaller finite sample performance in terms of bias though, because the OLS bias of the LEV structural equation is smaller than that of DIF, especially when the series are persistent. The weak instrument problem further manifests itself in poor performances of the Wald tests, which we show to have the same size distortions in the DIF and LEV models when $\sigma_\eta^2 = \sigma_v^2$. Although our theoretical results do not apply automatically to GMM based inference (Kiviet (2008)) we show by simulation that these properties generalise to the system GMM estimator.

Having established this potential weak instrument problem for the system GMM estimator, for inference one should therefore consider use of testing procedures that are robust to the weak instruments problem. The Kleibergen (2005) Lagrange Multiplier test and his GMM extension of the Conditional Likelihood Ratio test of Moreira (2003) are possible candidates, as is the Stock and Wright (2000) GMM version of the Anderson-Rubin statistic. Newey and Windmeijer (2009) show that the behaviours of these test statistics are not only robust to weak instrument asymptotics, they are also robust to many weak instrument asymptotics, where the number of instruments grow with the sample size, but with the model bounded away from non-identification. Newey and Windmeijer (2009) also propose use of the continuous updated GMM estimator (CUE, Hansen, Heaton and Yaron (1996)) with a new variance estimator that is valid under many weak instrument asymptotics. They show that the Wald test using the CUE estimation results and their proposed variance estimator performs well in a static panel data model estimated in first differences. As the number of potential instruments in this panel data setting grow quite rapidly with the time dimension of the panel, this may be a sensible approach also for the system moment conditions.

As a final remark, the direction of the biases of the DIF (downward) and LEV (upward) GMM estimators in the AR(1) panel data model are quite specific to this model specification. In different models these biases may be different and the SYS GMM estimator may have a larger absolute bias than the DIF GMM estimator. For example in the static panel data model

$$\begin{aligned} y_{it} &= x_{it}\beta + \eta_i + v_{it} \\ x_{it} &= \rho x_{i,t-1} + \gamma\eta_i + \delta v_{it} + w_{it} \end{aligned}$$

the DIF GMM estimator may have a smaller finite sample bias than the SYS GMM estimator when the x_{it} series are persistent, but $|\delta|$ is small and $|\gamma|$ is large, as then the endogeneity problem and OLS bias in the DIF model may be less than that of the LEV model.

Acknowledgements

We are grateful for helpful comments by Steve Bond, Jan Kiviet, Jon Temple and two anonymous referees.

References

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

[2] 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*, 277-298.

[3] Arellano, M. and O. Bover (1995). Another Look at the Instrumental Variable Estimation of Error-Components Models. *Journal of Econometrics 68*, 29-51.

[4] Blundell, R. and S. Bond (1998). Initial Conditions and Moment Restrictions in Dynamic Panel Data Models. *Journal of Econometrics 87*, 115-143.

[5] Blundell, R.W., S.R. Bond and F. Windmeijer (2000). Estimation in Dynamic Panel Data Models: Improving on the Performance of the Standard GMM Estimator. In B. Baltagi (Ed.), *Nonstationary Panels, Panel Cointegration, and Dynamic Panels, Advances in Econometrics 15*, 53-91, New York: JAI Press, Elsevier Science.

[6] Bond, S.R., A. Hoeffler and J. Temple (2001). GMM Estimation of Empirical Growth Models. Working Paper, University of Oxford.

[7] Bond, S.R. and F. Windmeijer (2005). Reliable Inference for GMM Estimators? Finite Sample Properties of Alternative Test Procedures in Linear Panel Data Models. *Econometric Reviews 24*, 1-37.

[8] Bound, J., D.A. Jaeger and R.M. Baker (1995). Problems with Instrumental Variables Estimation when the Correlation between the Instruments and the Endogenous Explanatory Variable is Weak. *Journal of the American Statistical Association 90*, 443-450.

[9] Bun, M.J.G. and J.F. Kiviet (2006). The Effects of Dynamic Feedbacks on LS and MM Estimator Accuracy in Panel Data Models. *Journal of Econometrics 132*, 409-444.

[10] Buse, A. (1992). The Bias of Instrumental Variable Estimators. *Econometrica 60*, 173-180.

[11] Griffith, R., R. Harrison and J. Van Reenen (2006). How Special is the Special Relationship? Using the Impact of U.S. R&D Spillovers on U.K. Firms as a Test of Technology Sourcing. *The American Economic Review 96*, 1859-1875.

[12] Hahn, J. and J. Hausman (2002). Note on Bias in Estimators for Simultaneous Equation Models. *Economics Letters 75*, 237-241.

[13] Hahn, J. and G. Kuersteiner (2002). Discontinuities of Weak Instrument Limiting Distributions. *Economics Letters 75*, 325-331.

[14] Han, C. and P.C.B. Phillips (2006). GMM with Many Moment Conditions. *Econometrica 74*, 147-192.

[15] Hansen, L.P., J. Heaton and A. Yaron (1996). Finite-Sample Properties of Some Alternative GMM Estimators. *Journal of Business and Economic Statistics 14*, 262-280.

[16] Hayakawa, K. (2007). Small Sample Bias Properties of the System GMM Estimator in Dynamic Panel Data Models. *Economics Letters 95*, 32-38.

[17] Hayakawa, K. (2008). The Asymptotic Properties of the System GMM Estimator in Dynamic Panel Data Models When Both N and T are Large. mimeo, *Hiroshima University.*

[18] Kiviet, J.F. (2007). Judging Contending Estimators by Simulation: Tournaments in Dynamic Panel Data Models. In G.D.A. Phillips and E. Tzavalis (Eds.), *The Refinement of Econometric Estimation and Test Procedures*, 282-318, Cambridge University Press.

[19] Kiviet, J.F. (2008). Strength and Weakness of Instruments in IV and GMM Estimation of Dynamic Panel Data Models. Working Paper, University of Amsterdam.

[20] Kleibergen, F. (2005). Testing Parameters in GMM Without Assuming They are Identified. *Econometrica 73*, 1103-1123.

[21] Kruiniger, H. (2009). GMM Estimation and Inference in Dynamic Panel Data Models with Persistent Data. Forhtcoming in *Econometric Theory.*

[22] Levine, R., N. Loayza and T. Beck (2000). Financial Intermediation and Growth: Causality and Causes. *Journal of Monetary Economics 46*, 31-77.

[23] Levinsohn, J. and A. Petrin (2003). Estimating Production Functions Using Inputs to Control for Unobservables. *Review of Economic Studies 70*, 317-341.

[24] Moreira, M. (2003). A Conditional Likelihood Ratio Test for Structural Models. *Econometrica 71*, 1027-1048.

[25] Morimune, K. (1989). t Test in a Structural Equation. *Econometrica 57*, 1341-1360.

[26] Nagar, A.L. (1959). The Bias and Moment Matrix of the General k-class Estimators of the parameters in Simultaneous Equations. *Econometrica 27*, 575-595.

[27] Newey, W.K. and F. Windmeijer (2009). GMM with Many Weak Moment Conditions. *Econometrica 77*, 687-719

[28] Picone, G.A., F. Sloan and J.G. Trogdon (2004). The Effect of the Tobacco Settlement and Smoking Bans on Alcohol Consumption. *Health Economics 13*, 1063-1080.

[29] Ridder, G. and T. Wansbeek (1990). Dynamic Models for Panel Data. In F. van der Ploeg (Ed.), *Advanced Lectures in Quantitative Economics*, 557-582. London: Academic Press.

[30] Rothenberg, T.J. (1984). Approximating the Distributions of Econometric Estimators and Test Statistics. In Z. Griliches and M.D. Intriligator (Eds.), *Handbook of Econometrics, Volume 2*, 881-935. Amsterdam: North Holland.

[31] Staiger, D. and J.H. Stock (1997). Instrumental Variables Regression with Weak Instruments. *Econometrica 65*, 557-586.

[32] Stock, J.H. and J.H. Wright (2000). GMM with Weak Identification. *Econometrica 68*, 1055-1096.

[33] Stock, J.H., J.H. Wright and M. Yogo (2002). A Survey of Weak Instruments and Weak Identification in Generalized Method of Moments. *Journal of Business & Economic Statistics 20*, 518-529.

[34] Stock, J.H. and M. Yogo (2005). Testing for Weak Instruments in Linear IV Regression. In D.W.K. Andrews and J.H. Stock (Eds.), *Identification and Inference for Econometric Models, Essays in Honor of Thomas Rothenberg*, 80-108. New York: Cambridge University Press.

[35] Windmeijer, F. (2005). A Finite Sample Correction for the variance of Linear Efficient Two-Step GMM Estimators. *Journal of Econometrics 126*, 25-517.

Appendix

A.1 Concentration Parameters in Cross-Section Analysis

The model in first differences for the cross-section at time t is given by

$$\begin{aligned} \Delta y_{it} &= \alpha \Delta y_{i,t-1} + \Delta u_{it} \\ \Delta y_{i,t-1} &= y_i^{t-2\prime} \pi_{dt} + d_{i,t-1}. \end{aligned}$$

For the general expression of the expected value of the concentration parameter divided by n we get

$$E\left(\frac{1}{n}\mu_{dt}\right) = \frac{\pi_{dt}' E\left(y_i^{t-2} y_i^{t-2\prime}\right) \pi_{dt}}{\sigma_{dt}^2}$$

but as

$$\pi_{dt} = \left[E\left(y_i^{t-2} y_i^{t-2\prime}\right)\right]^{-1} E\left(y_i^{t-2} \Delta y_{i,t-1}\right)$$

and

$$\sigma_{dt}^2 = E\left(\left(\Delta y_{i,t-1} - y_i^{t-2\prime} \pi_{dt}\right)^2\right)$$

we get

$$E\left(\frac{1}{n}\mu_{dt}\right) = \frac{\left(E\left(y_i^{t-2}\Delta y_{i,t-1}\right)\right)' \left[E\left(y_i^{t-2} y_i^{t-2\prime}\right)\right]^{-1} E\left(y_i^{t-2}\Delta y_{i,t-1}\right)}{E\left(\Delta y_{i,t-1}^2\right) - \left(E\left(y_i^{t-2}\Delta y_{i,t-1}\right)\right)' \left[E\left(y_i^{t-2} y_i^{t-2\prime}\right)\right]^{-1} E\left(y_i^{t-2}\Delta y_{i,t-1}\right)}.$$

Under covariance stationarity

$$E\left(y_i^{t-2} y_i^{t-2\prime}\right) = \frac{\sigma_\eta^2}{(1-\alpha)^2}\iota_{t-2}\iota_{t-2}' + \frac{\sigma_v^2}{1-\alpha^2} G_{t-2}$$

where

$$G_{t-2} = \begin{bmatrix} 1 & \alpha & \cdots & \alpha^{t-3} \\ \alpha & 1 & & \vdots \\ \vdots & & \ddots & \alpha \\ \alpha^{t-3} & \cdots & \alpha & 1 \end{bmatrix}.$$

The inverse of $E\left(y_i^{t-2} y_i^{t-2\prime}\right)$ is given by (see e.g. Ridder and Wansbeek (1990))

$$\left[E\left(y_i^{t-2} y_i^{t-2\prime}\right)\right]^{-1} = \frac{1}{\sigma_v^2}\left[R_{t-2}' R_{t-2} - \frac{\sigma_\eta^2 h_{t-2} h_{t-2}'}{\sigma_v^2 + \sigma_\eta^2\left(t-3+\frac{1+\alpha}{1-\alpha}\right)}\right]$$

where

$$R_{t-2} = \begin{bmatrix} 1 & -\alpha & 0 & & 0 \\ 0 & 1 & -\alpha & & \\ & & \ddots & \ddots & \\ & & & 1 & -\alpha \\ 0 & & & 0 & \sqrt{1-\alpha^2} \end{bmatrix}; \; h_{t-2} = (1-\alpha)\,\iota_{t-2} + \alpha\,(e_1 + e_{t-2})$$

and e_j is the j-th unit vector of order $t-2$.

We further have that

$$E\left(y_i^{t-2}\Delta y_{i,t-1}\right) = -\frac{\sigma_v^2}{1+\alpha} g_{t-2}$$

where

$$g_{t-2} = \begin{bmatrix} \alpha^{t-3} \\ \vdots \\ \alpha \\ 1 \end{bmatrix}.$$

As

$$R_{t-2}g_{t-2} = \begin{bmatrix} 0 \\ \vdots \\ 0 \\ \sqrt{1-\alpha^2} \end{bmatrix}; \; h'_{t-2}g_{t-2} = 1+\alpha$$

and so

$$\begin{aligned} & \left(E\left(y_i^{t-1}\Delta y_{i,t-1}\right)\right)' \left[E\left(y_i^{t-2}y_i^{t-2\prime}\right)\right]^{-1} E\left(y_i^{t-2}\Delta y_{i,t-1}\right) \\ = \; & \frac{\sigma_v^2}{(1+\alpha)^2}\left(1-\alpha^2 - \frac{\sigma_\eta^2(1+\alpha)^2}{\sigma_v^2 + \sigma_\eta^2\left(t-3+\frac{1+\alpha}{1-\alpha}\right)}\right). \end{aligned}$$

Further

$$E\left(\Delta y_{i,t-1}^2\right) = \frac{2\sigma_v^2}{1+\alpha}.$$

Combining these results in

$$
\begin{aligned}
E\left(\frac{1}{n}\mu_{dt}\right) &= \frac{\frac{\sigma_v^2}{(1+\alpha)^2}\left(1-\alpha^2-\frac{\sigma_\eta^2(1+\alpha)^2}{\sigma_v^2+\sigma_\eta^2\left(t-3+\frac{1+\alpha}{1-\alpha}\right)}\right)}{\frac{2\sigma_v^2}{1+\alpha}-\frac{\sigma_v^2}{(1+\alpha)^2}\left(1-\alpha^2-\frac{\sigma_\eta^2(1+\alpha)^2}{\sigma_v^2+\sigma_\eta^2\left(t-3+\frac{1+\alpha}{1-\alpha}\right)}\right)} \\
&= \frac{1-\alpha^2-\frac{\sigma_\eta^2(1+\alpha)^2}{\sigma_v^2+\sigma_\eta^2\left(t-3+\frac{1+\alpha}{1-\alpha}\right)}}{2\left(1+\alpha\right)-\left(1-\alpha^2-\frac{\sigma_\eta^2(1+\alpha)^2}{\sigma_v^2+\sigma_\eta^2\left(t-3+\frac{1+\alpha}{1-\alpha}\right)}\right)} \\
&= \frac{\left(1-\alpha^2\right)\left(\sigma_v^2+\sigma_\eta^2\left(t-3+\frac{1+\alpha}{1-\alpha}\right)\right)-\sigma_\eta^2\left(1+\alpha\right)^2}{\left(1+\alpha\right)^2\left(\sigma_v^2+\sigma_\eta^2\left(t-3+\frac{1+\alpha}{1-\alpha}\right)\right)+\sigma_\eta^2\left(1+\alpha\right)^2} \\
&= \frac{\left(1-\alpha\right)\left(\sigma_v^2+\left(t-3\right)\sigma_\eta^2\right)}{\left(1+\alpha\right)\left(\sigma_v^2+\sigma_\eta^2\left(t-2+\frac{1+\alpha}{1-\alpha}\right)\right)} \\
&= \frac{\left(1-\alpha\right)^2\left(\sigma_v^2+\left(t-3\right)\sigma_\eta^2\right)}{\left(1-\alpha^2\right)\sigma_v^2+\left(\left(t-1\right)-\left(t-3\right)\alpha\right)\left(1+\alpha\right)\sigma_\eta^2}.
\end{aligned}
$$

For the model in levels we have for the cross-section at time t

$$
\begin{aligned}
y_{it} &= \alpha y_{i,t-1}+\eta_i+v_{it} \\
y_{i,t-1} &= \Delta y_i^{t-1\prime}\pi_{lt}+l_{i,t-1}
\end{aligned}
$$

and the expected concentration parameter is given by

$$
E\left(\frac{1}{n}\mu_{lt}\right)=\frac{\left(E\left(y_{i,t-1}\Delta y_i^{t-1}\right)\right)'\left[E\left(\Delta y_i^{t-1}\Delta y_i^{t-1\prime}\right)\right]^{-1}E\left(y_{i,t-1}\Delta y_i^{t-1}\right)}{E\left(y_{i,t-1}^2\right)-\left(E\left(y_{i,t-1}\Delta y_i^{t-1}\right)\right)'\left[E\left(\Delta y_i^{t-1}\Delta y_i^{t-1\prime}\right)\right]^{-1}E\left(y_{i,t-1}\Delta y_i^{t-1}\right)}.
$$

Again, under covariance stationarity, we have that

$$
E\left(\Delta y_i^{t-1}\Delta y_i^{t-1\prime}\right)=\frac{\sigma_v^2}{1+\alpha}\begin{bmatrix} 2 & \alpha-1 & \alpha\left(\alpha-1\right) & \cdots & \alpha^{t-4}\left(\alpha-1\right) \\ \alpha-1 & 2 & \alpha-1 & \cdots & \alpha^{t-5}\left(\alpha-1\right) \\ \alpha\left(\alpha-1\right) & \alpha-1 & 2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \alpha-1 \\ \alpha^{t-4}\left(\alpha-1\right) & \cdots & \alpha\left(\alpha-1\right) & \alpha-1 & 2 \end{bmatrix}
$$

and

$$
E\left(y_{i,t-1}\Delta y_i^{t-1}\right)=\frac{\sigma_v^2}{1+\alpha}\begin{bmatrix} \alpha^{t-3} \\ \vdots \\ \alpha \\ 1 \end{bmatrix}.
$$

It then follows that

$$E\left(y_{i,t-1}\Delta y_i^{t-1}\right)'\left[E\left(\Delta y_i^{t-1}\Delta y_i^{t-1\prime}\right)\right]^{-1}E\left(y_{i,t-1}\Delta y_i^{t-1}\right)=\frac{(t-2)\,\sigma_v^2}{(1+\alpha)\left((t-1)-(t-3)\,\alpha\right)}.$$

As

$$E\left(y_{i,t-1}^2\right)=\frac{\sigma_\eta^2}{(1-\alpha)^2}+\frac{\sigma_v^2}{1-\alpha^2}$$

we get that

$$\begin{aligned}
E\left(\frac{1}{n}\mu_{lt}\right) &= \frac{\frac{(t-2)\sigma_v^2}{(1+\alpha)((t-1)-(t-3)\alpha)}}{\frac{\sigma_\eta^2}{(1-\alpha)^2}+\frac{\sigma_v^2}{1-\alpha^2}-\frac{(t-2)\sigma_v^2}{(1+\alpha)((t-1)-(t-3)\alpha)}} \\
&= \frac{(t-2)\,\sigma_v^2}{(1+\alpha)\left((t-1)-(t-3)\,\alpha\right)\left(\frac{\sigma_\eta^2}{(1-\alpha)^2}+\frac{\sigma_v^2}{1-\alpha^2}\right)-(t-2)\,\sigma_v^2} \\
&= \frac{(t-2)\,(1-\alpha)^2\,\sigma_v^2}{\left((t-1)-(t-3)\,\alpha\right)\left((1+\alpha)\,\sigma_\eta^2+(1-\alpha)\,\sigma_v^2\right)-(t-2)\,(1-\alpha)^2\,\sigma_v^2} \\
&= \frac{(t-2)\,(1-\alpha)^2\,\sigma_v^2}{\left(1-\alpha^2\right)\sigma_v^2+\left((t-1)-(t-3)\,\alpha\right)(1+\alpha)\,\sigma_\eta^2}.
\end{aligned}$$

A.2 Mean Stationarity Only

We now relax the assumption of covariance stationarity, while maintaining mean stationarity, i.e. we specify the initial condition as

$$y_{i1}=\frac{\eta_i}{1-\alpha}+\varepsilon_i$$

with $E\left(\varepsilon_i^2\right)=\sigma_\varepsilon^2$.

For $t=3$, we get in this case

$$\pi_{d3}=\frac{E\left(y_1\Delta y_2\right)}{E\left(y_1^2\right)}=-\frac{(1-\alpha)\,\sigma_\varepsilon^2}{\frac{\sigma_\eta^2}{(1-\alpha)^2}+\sigma_\varepsilon^2}=-\frac{(1-\alpha)\,\sigma_\varepsilon^2}{\sigma_{y_1}^2}$$

$$\begin{aligned}
\sigma_{d3}^2 &= E\left(\Delta y_2\right)^2-2\pi_{d3}E\left(y_1\Delta y_2\right)+\pi_{d3}^2E\left(y_1^2\right) \\
&= \sigma_v^2+(1-\alpha)^2\,\sigma_\varepsilon^2+\pi_{d3}\,(1-\alpha)\,\sigma_\varepsilon^2
\end{aligned}$$

$$\begin{aligned}\mu_{d3} &= \frac{\pi_{d3}^2 y_1' y_1}{\sigma_{d3}^2} \\ &= \frac{\pi_{d3}^2}{\sigma_v^2 + (1-\alpha)^2 \sigma_\varepsilon^2 + \pi_{d3}(1-\alpha)\sigma_\varepsilon^2} y_1' y_1.\end{aligned}$$

$$\begin{aligned}E\left(\frac{1}{n}\mu_{d3}\right) &= \frac{\left(\frac{(1-\alpha)\sigma_\varepsilon^2}{\sigma_{y_1}^2}\right)^2}{\sigma_v^2 + (1-\alpha)^2 \sigma_\varepsilon^2 - \frac{((1-\alpha)\sigma_\varepsilon^2)^2}{\sigma_{y_1}^2}} \sigma_{y_1}^2 \\ &= \frac{\frac{\left((1-\alpha)\sigma_\varepsilon^2\right)^2}{\sigma_{y_1}^2}}{\sigma_v^2 + (1-\alpha)^2 \sigma_\varepsilon^2 - \frac{((1-\alpha)\sigma_\varepsilon^2)^2}{\sigma_{y_1}^2}}\end{aligned}$$

For the levels model we get

$$\begin{aligned}\pi_{l3} &= \frac{E(y_2 \Delta y_2)}{E\left((\Delta y_2)^2\right)} \\ &= \frac{\sigma_v^2 - \alpha(1-\alpha)\sigma_\varepsilon^2}{\sigma_v^2 + (1-\alpha)^2 \sigma_\varepsilon^2}\end{aligned}$$

and

$$\begin{aligned}\sigma_{l3}^2 &= E\left(y_2^2\right) - \pi_{l3} E(y_2 \Delta y_2) \\ &= \frac{\sigma_\eta^2}{(1-\alpha)^2} + \sigma_v^2 + \alpha^2 \sigma_\varepsilon^2 - \frac{\left(\sigma_v^2 - \alpha(1-\alpha)\sigma_\varepsilon^2\right)^2}{\sigma_v^2 + (1-\alpha)^2 \sigma_\varepsilon^2}.\end{aligned}$$

The concentration parameter is therefore given by

$$\begin{aligned}\mu_{l3} &= \frac{\pi_{l3}^2 \Delta y_2' \Delta y_2}{\sigma_{l3}^2} \\ &= \frac{\left(\frac{\sigma_v^2 - \alpha(1-\alpha)\sigma_\varepsilon^2}{\sigma_v^2 + (1-\alpha)^2\sigma_\varepsilon^2}\right)^2}{\frac{\sigma_\eta^2}{(1-\alpha)^2} + \sigma_v^2 + \alpha^2\sigma_\varepsilon^2 - \frac{(\sigma_v^2 - \alpha(1-\alpha)\sigma_\varepsilon^2)^2}{\sigma_v^2 + (1-\alpha)^2\sigma_\varepsilon^2}} \Delta y_2' \Delta y_2\end{aligned}$$

and so

$$E\left(\frac{1}{n}\mu_{l3}\right) = \frac{\frac{\left(\sigma_v^2 - \alpha(1-\alpha)\sigma_\varepsilon^2\right)^2}{\sigma_v^2 + (1-\alpha)^2\sigma_\varepsilon^2}}{\frac{\sigma_\eta^2}{(1-\alpha)^2} + \sigma_v^2 + \alpha^2\sigma_\varepsilon^2 - \frac{(\sigma_v^2 - \alpha(1-\alpha)\sigma_\varepsilon^2)^2}{\sigma_v^2 + (1-\alpha)^2\sigma_\varepsilon^2}}.$$

Calculating these expectations shows that $E\left(\frac{1}{n}\mu_{l3}\right) > E\left(\frac{1}{n}\mu_{d3}\right)$ if $\sigma_\varepsilon^2 < \frac{\sigma_v^2}{1-\alpha^2}$ and $E\left(\frac{1}{n}\mu_{l3}\right) < E\left(\frac{1}{n}\mu_{d3}\right)$ if $\sigma_\varepsilon^2 > \frac{\sigma_v^2}{1-\alpha^2}$, i.e. the expected concentration parameter in the levels model is larger than that of the differenced model if the variance of the initial condition is smaller than the covariance stationary level and vice versa.

A.3 Bias Approximations for Panel 2SLS Estimators

We will first evaluate the bias approximation for the panel DIF estimator. Note that due to the block-diagonal structure of the Z_{di} instrument matrix we have

$$\left(Z_d'Z_d\right)^{-1} = diag\left[\left(Z_{d3}'Z_{d3}\right)^{-1}, ..., \left(Z_{dT}'Z_{dT}\right)^{-1}\right],$$

where the $n \times (t-2)$ matrix Z_{dt} is $y^{t-2} = (y_1^{t-2}, ..., y_n^{t-2})'$. Hence, we can write

$$\Delta y_{-1}' Z_d \left(Z_d'Z_d\right)^{-1} Z_d'\Delta y_{-1} = \sum_{t=3}^{T} \Delta y_{t-1}' Z_{dt} \left(Z_{dt}'Z_{dt}\right)^{-1} Z_{dt}'\Delta y_{t-1},$$

$$\Delta y_{-1}' Z_d \left(Z_d'Z_d\right)^{-1} Z_d'\Delta u = \sum_{t=3}^{T} \Delta y_{t-1}' Z_{dt} \left(Z_{dt}'Z_{dt}\right)^{-1} Z_{dt}'\Delta u_t.$$

Exploiting $\Delta y_{t-1} = Z_{dt}\pi_{dt} + d_{t-1}$ and defining $P_{dt} = Z_{dt}\left(Z_{dt}'Z_{dt}\right)^{-1} Z_{dt}'$ we have

$$\begin{aligned} E\left(\Delta y_{t-1}' Z_{dt} \left(Z_{dt}'Z_{dt}\right)^{-1} Z_{dt}'\Delta y_{t-1}\right) &= \pi_{dt}' E\left(Z_{dt}'Z_{dt}\right)\pi_{dt} + E\left(d_{t-1}' P_{dt} d_{t-1}\right) \\ &= \sigma_{dt}^2 (E\left(\mu_{dt}\right) + (t-2)). \end{aligned}$$

The expectation of the numerator of the estimation error is

$$E\left(\Delta y_{t-1}' Z_{dt} \left(Z_{dt}'Z_{dt}\right)^{-1} Z_{dt}'\Delta u_t\right) = \sigma_{\Delta u,d}(t-2).$$

Combining results we have

$$\begin{aligned} E\left(\hat{\alpha}_d - \alpha\right) &\approx \frac{\sum_{t=3}^{T} \sigma_{\Delta u,d}(t-2)}{\sum_{t=3}^{T} \sigma_{dt}^2 (E\left(\mu_{dt}\right) + (t-2))} \\ &= \frac{0.5(T-1)(T-2)\sigma_{\Delta u,d}}{\sum_{t=3}^{T} \sigma_{dt}^2 (E\left(\mu_{dt}\right) + (t-2))}. \end{aligned}$$

The bias approximation for the panel LEV estimator can be derived in the same way. Regarding the SYS estimator we can write

$$q_{-1}'Z_s\left(Z_s'Z_s\right)^{-1}Z_s'q_{-1} = \sum_{t=3}^{T}\Delta y_{t-1}'Z_{dt}\left(Z_{dt}'Z_{dt}\right)^{-1}Z_{dt}'\Delta y_{t-1}+\sum_{t=3}^{T}y_{t-1}'Z_{lt}\left(Z_{lt}'Z_{lt}\right)^{-1}Z_{lt}'y_{t-1},$$

$$q_{-1}'Z_s\left(Z_s'Z_s\right)^{-1}Z_s'p = \sum_{t=3}^{T}\Delta y_{t-1}'Z_{dt}\left(Z_{dt}'Z_{dt}\right)^{-1}Z_{dt}'\Delta u_t+\sum_{t=3}^{T}y_{t-1}'Z_{lt}\left(Z_{lt}'Z_{lt}\right)^{-1}Z_{lt}'u_t.$$

It should be noted that only the non-redundant LEV moment conditions have been used in system estimation. In other words, Z_{li} and, hence, Z_{lt} in system estimation are defined as

$$Z_{li}=\begin{bmatrix}\Delta y_{i2} & 0 & \cdots & 0\\ 0 & \Delta y_{i3} & \cdots & 0\\ . & . & \cdots & .\\ 0 & 0 & \cdots & \Delta y_{iT-1}\end{bmatrix},\quad Z_{lt}=\begin{bmatrix}\Delta y_{1,t-1}\\ \Delta y_{2,t-1}\\ .\\ \Delta y_{n,t-1}\end{bmatrix},$$

hence we exploit one instrument per period only. As a result we have

$$\begin{aligned}E\left(y_{t-1}'Z_{lt}\left(Z_{lt}'Z_{lt}\right)^{-1}Z_{lt}'y_{t-1}\right) &= \pi_{lt}'E\left(Z_{lt}'Z_{lt}\right)\pi_{lt}+E\left(l_{t-1}'P_{lt}l_{t-1}\right)\\ &= \sigma_{lt}^2(E\left(\mu_{lt}\right)+1).\end{aligned}$$

and

$$E\left(\Delta y_{t-1}'Z_{lt}\left(Z_{lt}'Z_{lt}\right)^{-1}Z_{lt}'u_t\right)=\sigma_{u,l}.$$

Combining results we find

$$\begin{aligned}E\left(\hat{\alpha}_s-\alpha\right) &\approx \frac{E\left(\sum_{t=3}^{T}\Delta y_{t-1}'Z_{dt}\left(Z_{dt}'Z_{dt}\right)^{-1}Z_{dt}'\Delta u_t+\sum_{t=3}^{T}y_{t-1}'Z_{lt}\left(Z_{lt}'Z_{lt}\right)^{-1}Z_{lt}'u_t\right)}{E\left(\sum_{t=3}^{T}\Delta y_{t-1}'Z_{dt}\left(Z_{dt}'Z_{dt}\right)^{-1}Z_{dt}'\Delta y_{t-1}+\sum_{t=3}^{T}y_{t-1}'Z_{lt}\left(Z_{lt}'Z_{lt}\right)^{-1}Z_{lt}'y_{t-1}\right)}\\ &= \frac{0.5(T-1)(T-2)\sigma_{\Delta u,d}+(T-2)\sigma_{u,l}}{\sum_{t=3}^{T}\sigma_{dt}^2(E\left(\mu_{dt}\right)+(t-2))+\sum_{t=3}^{T}\sigma_{lt}^2(E\left(\mu_{lt}\right)+1)}.\end{aligned}$$