

# THE WEAK INSTRUMENT PROBLEM OF THE SYSTEM GMM ESTIMATOR IN DYNAMIC PANEL DATA MODELS

---

*Maurice Bun  
Frank Windmeijer*

THE INSTITUTE FOR FISCAL STUDIES  
DEPARTMENT OF ECONOMICS, UCL  
cemmap working paper CWP08/07

# 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

March, 2007

## 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 in a Monte Carlo study to extend to the panel data system GMM estimator.

**JEL Classification:** C12, C13, C23

**Keywords:** Dynamic Panel Data, System GMM, Weak Instruments

---

\*Helpful comments were provided by Steve Bond and Jon Temple. This paper has been presented at the 13th Panel Data Conference in Cambridge, Amsterdam, Copenhagen, cemmap and Bristol.

# 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. 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 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 ( $\sigma_{\eta}^2$ ) is equal to the variance of the idiosyncratic shocks ( $\sigma_v^2$ ). 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 Stock and Yogo (2005), 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. Therefore the absolute bias of the LEV 2SLS estimator is smaller than that of the DIF 2SLS estimator when the series are persistent.

Results in Stock and Yogo (2005) further indicate that we 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  $\sigma_\eta^2$  is smaller than  $\sigma_v^2$ , 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  $\sigma_\eta^2$  is larger than  $\sigma_v^2$ . Also, unlike for DIF, the LEV OLS bias increases with increasing  $\sigma_\eta^2/\sigma_v^2$  and therefore the performances of the LEV and SYS 2SLS estimators deteriorate with increasing  $\sigma_\eta^2$ . 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 (2005), and explain 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  $\sigma_\eta^2/\sigma_v^2$  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  $\sigma_\eta^2/\sigma_v^2$ . 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 results on (relative) biases and Wald tests size distortions for the 2SLS estimators. Section 5 presents Monte Carlo results for the GMM panel data estimators. Section 6 concludes.

## 2 Model and GMM Estimators

We consider the first-order autoregressive panel data model

$$\begin{aligned} y_{it} &= \alpha y_{i,t-1} + u_{it}, & i = 1, \dots, N; t = 2, \dots, T, \\ u_{it} &= \eta_i + v_{it} \end{aligned} \tag{1}$$

where it is assumed that  $\eta_i$  and  $v_{it}$  have an error components structure with

$$E(\eta_i) = 0, \quad E(v_{it}) = 0, \quad E(v_{it}\eta_i) = 0, \quad i = 1, \dots, N; t = 2, \dots, T \tag{2}$$

$$E(v_{it}v_{is}) = 0, \quad i = 1, \dots, N \text{ and } t \neq s, \tag{3}$$

and the initial condition satisfies

$$E(y_{i1}v_{it}) = 0, \quad i = 1, \dots, N; t = 2, \dots, T. \tag{4}$$

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

$$E(y_i^{t-2} \Delta u_{it}) = 0, \quad t = 3, \dots, T, \quad (5)$$

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

Defining

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

moment conditions (5) can be more compactly written as

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

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

$$\hat{\alpha}_d = \frac{\Delta y'_{-1} Z_d W_N^{-1} Z'_d \Delta y}{\Delta y'_{-1} Z_d W_N^{-1} Z'_d \Delta y_{-1}}$$

where  $\Delta y = (\Delta y'_1, \Delta y'_2, \dots, \Delta y'_N)'$ ,  $\Delta y_i = (\Delta y_{i3}, \Delta y_{i4}, \dots, \Delta y_{iT})'$ ,  $\Delta y_{-1}$  the lagged version of  $\Delta y$ ,  $Z_d = (Z'_{d1}, Z'_{d2}, \dots, Z'_{dN})'$  and  $W_N$  is a weight matrix determining the efficiency properties of the GMM estimator. Clearly,  $\hat{\alpha}_d$  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

$$E(\eta_i \Delta y_{i2}) = 0, \quad (7)$$

which holds when the process is mean stationary:

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

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

$$E(u_{it} \Delta y_i^{t-1}) = 0, \quad t = 3, \dots, T, \quad (9)$$

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

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

moment conditions (9) can be written as

$$E(Z_{li}' u_i) = 0, \quad (10)$$

with the GMM estimator based on these moment conditions given by

$$\hat{\alpha}_l = \frac{y_{-1}' Z_l W_N^{-1} Z_l' y}{y_{-1}' Z_l W_N^{-1} Z_l' y_{-1}},$$

where we will refer to  $\hat{\alpha}_l$  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

$$E(y_i^{t-2} \Delta u_{it}) = 0 \quad t = 3, \dots, T; \quad (11)$$

$$E(u_{it} \Delta y_{i,t-1}) = 0 \quad t = 3, \dots, T,$$

or

$$E(Z_{si}' p_i) = 0, \quad (12)$$

where

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

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  $E(\varepsilon_i^2) = \frac{\sigma_u^2}{1-\alpha^2}$  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  $i = 1, \dots, N$ , where the  $(u_i, \varepsilon_i)$  are independent draws from a bivariate normal distribution with zero means, variances  $\sigma_u^2$  and  $\sigma_\varepsilon^2$ , and correlation coefficient  $\rho$ . The parameter  $\beta$  is estimated by 2SLS:

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

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

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) and Stock, Wright and Yogo (2002). With weak instruments, the 2SLS estimator is biased in the direction of the OLS estimator, and its distribution 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

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

it is clear that  $\hat{\mu}$  is equal to the Wald test for testing the hypothesis  $H_0 : \pi = 0$ , and  $\hat{\mu}/k_z$  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

$$\hat{\beta} = \beta + \frac{\pi' Z' u + \xi' P_Z u}{\pi' Z' Z \pi + 2\pi' Z' \xi + \xi' P_Z \xi}, \quad (14)$$

and so

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

where

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

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

$(A, B)$  is bivariate normal with zero means, unit variances and correlation coefficient  $\rho$ .  $s$  has mean  $k_z \rho$  and variance  $k_z (1 + \rho^2)$  and  $S$  has mean  $k_z$  and variance  $2k_z$ . It is clear that when  $\mu$  is large,  $\sqrt{\mu} (\hat{\beta} - \beta)$  behaves like the  $N(0, 1)$  random variable  $B$ .

Using weak instrument asymptotics, Stock and Yogo (2005) tabulate critical values for the first-stage F-statistic to test whether given instruments are weak. They do this separately for the maximum bias of the IV estimator, relative to the bias of the OLS estimator, and for the maximum Wald test size distortion.

## 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

$$\begin{aligned}\Delta y_{it} &= \alpha \Delta y_{i,t-1} + \Delta u_{it} \\ \Delta y_{i,t-1} &= y_i^{t-2'} \pi_{dt} + d_{i,t-1}^t.\end{aligned}$$

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

$$E\left(\frac{1}{N}\mu_{dt}\right) = \frac{\pi'_{dt} E(y_i^{t-2} y_i^{t-2'}) \pi_{dt}}{\sigma_{dt}^2}.$$

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

$$\begin{aligned}y_{it} &= \alpha y_{i,t-1} + \eta_i + v_{it} \\ y_{i,t-1} &= \Delta y_i^{t-1'} \pi_{lt} + l_{i,t-1}^t\end{aligned}$$

and the expected concentration parameter is given by

$$E\left(\frac{1}{N}\mu_{lt}\right) = \frac{\pi'_{lt} E(\Delta y_i^{t-1} \Delta y_i^{t-1'}) \pi_{lt}}{\sigma_{lt}^2}.$$

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

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

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

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

Therefore

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

and for  $t > 3$

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

Figure 1 graphs the values of  $E\left(\frac{1}{N}\mu_{dt}\right)$  and  $E\left(\frac{1}{N}\mu_{dl}\right)$  as a function of  $\alpha$  for  $t = 6$  and various values of  $\frac{\sigma_\eta^2}{\sigma_v^2} = \{\frac{1}{4}, 1, 4\}$ . The values of the concentration parameters decrease with increasing  $\alpha$ . 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.

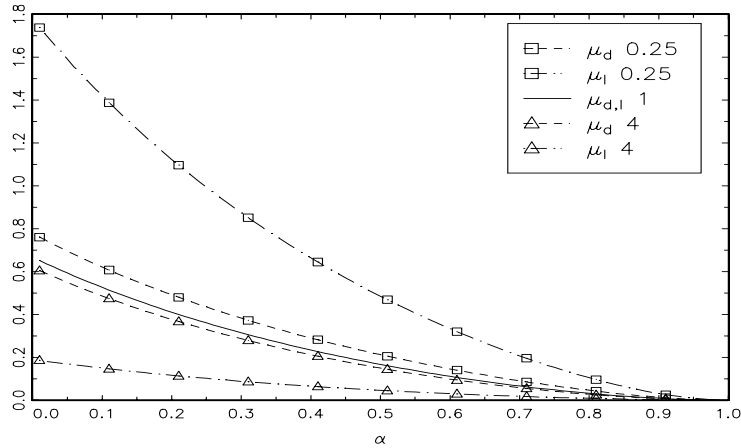


Figure 1.  $E\left(\frac{1}{N}\mu\right)$  as a function of  $\alpha$ ,  $t = 6$  and  $\frac{\sigma_\eta^2}{\sigma_v^2} = \{\frac{1}{4}, 1, 4\}$ .

## 4.1 Discussion

The fact that the concentration parameters are the same 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  $\sigma_\eta^2 = \sigma_v^2 = 1$ ,

and where  $\hat{\alpha}_l$  outperforms  $\hat{\alpha}_d$  in terms of bias and rmse, especially when the series become more persistent, i.e. when  $\alpha$  gets larger. The identification problem is apparent in the DIF model, where the reduced form parameters approach zero when  $\alpha$  approaches 1. This is in sharp contrast to the reduced form parameters in the LEV model that approach  $\frac{1}{2}$  when  $\alpha$  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  $\alpha$  for larger values of  $\alpha$ .

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 and therefore different biases of the OLS estimator 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

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

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

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

For the LEV model

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

the OLS estimator is given by

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

and the limiting bias of the OLS estimator is given by

$$\text{plim}(\hat{\alpha}_{iOLS} - \alpha) = (1 - \alpha) \frac{\frac{\sigma_\eta^2}{\sigma_v^2}}{\frac{\sigma_\eta^2}{\sigma_v^2} + \frac{1-\alpha}{1+\alpha}}$$

which reduces to  $\text{plim}(\hat{\alpha}_{iOLS} - \alpha) = (1 - \alpha^2)/2$  when  $\sigma_\eta^2 = \sigma_v^2$ .

The asymptotic absolute bias of  $\hat{\alpha}_{iOLS}$  is therefore (much) smaller than that of  $\hat{\alpha}_{dOLS}$  for high values of  $\alpha$ . Stock and Yogo (2005) relate the value of the concentration parameter to the absolute bias of the 2SLS estimator, relative to the absolute bias of the OLS estimator. When the concentration parameters are the same, we expect therefore that the relative biases are the same for the DIF and LEV 2SLS estimators. But the absolute bias of the LEV 2SLS estimator will then be smaller than that of the DIF estimator. From the results of Stock and Yogo (2005) we further expect the Wald test statistics to behave similarly when testing parameter restrictions. When the concentration parameters are small there will be significant size distortions.

## 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"

$$\begin{pmatrix} \Delta y_t \\ y_t \end{pmatrix} = \alpha \begin{pmatrix} \Delta y_{t-1} \\ y_{t-1} \end{pmatrix} + \begin{pmatrix} \Delta u_t \\ u_t \end{pmatrix} \quad (15)$$

is given by

$$\hat{\alpha}_{sOLS} = (\Delta y_t' \Delta y_t + y_{t-1}' y_{t-1})^{-1} (\Delta y_{t-1}' \Delta y_t + y_{t-1}' y_t)$$

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

$$\hat{\alpha}_{sOLS} = \tilde{\gamma} \hat{\alpha}_{dOLS} + (1 - \tilde{\gamma}) \hat{\alpha}_{iOLS}$$

where

$$\tilde{\gamma} = \frac{\Delta y_t' \Delta y_t}{\Delta y_t' \Delta y_t + y_{t-1}' y_{t-1}}$$

and

$$\text{plim } \tilde{\gamma} = \frac{1 - \alpha}{2 + \frac{1}{2} \frac{\sigma_\eta^2}{\sigma_v^2} \frac{1 + \alpha}{1 - \alpha}}.$$

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

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

Figure 2 shows the asymptotic biases of the DIF, LEV and SYS OLS estimators as a function of  $\alpha$  for different values of  $\sigma_\eta^2/\sigma_v^2 = \{\frac{1}{4}, 1, 4\}$ . 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  $\alpha$ .

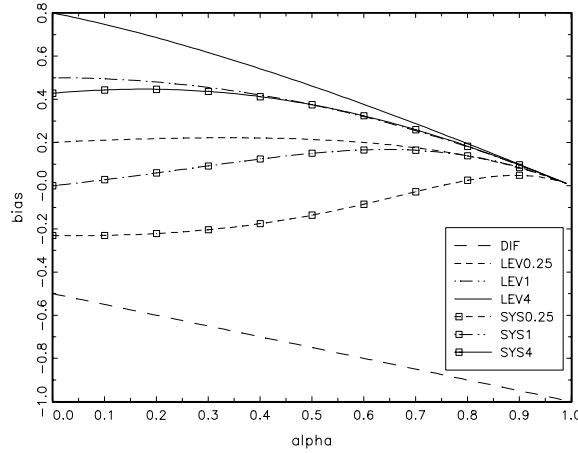


Figure 2. Asymptotic biases of OLS estimators,  $\sigma_\eta^2/\sigma_v^2 = \{\frac{1}{4}, 1, 4\}$ .

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

$$\hat{\alpha}_s = \tilde{\delta} \hat{\alpha}_d + (1 - \tilde{\delta}) \hat{\alpha}_l$$

where

$$\tilde{\delta} = \frac{\hat{\pi}'_d Z'_d Z_d \hat{\pi}_d}{\hat{\pi}'_d Z'_d Z_d \hat{\pi}_d + \hat{\pi}'_l Z'_l Z_l \hat{\pi}_l},$$

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

$$\text{plim } \tilde{\delta} = \frac{E\left(\frac{1}{N}\mu_d\right)}{E\left(\frac{1}{N}\mu_d\right) + \frac{\sigma_\eta^2}{\sigma_d^2}E\left(\frac{1}{N}\mu_l\right)}$$

and again  $\tilde{\delta} \rightarrow 0$  if  $\alpha \rightarrow 1$  and/or  $\sigma_\eta^2/\sigma_v^2 \rightarrow \infty$ . 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  $\alpha$ , 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  $(1 - \tilde{\delta})$  an increasing function in  $\frac{\sigma_\eta^2}{\sigma_v^2}$ , implying that more weight will be given to the LEV moment conditions.

Clearly, the SYS 2SLS estimator is not efficient as there is heteroskedasticity and correlation between the errors in model (15). 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

$$\begin{aligned} y_{i1} &= \frac{\eta_i}{1 - \alpha} + \varepsilon_i; \\ y_{it} &= \alpha y_{i,t-1} + \eta_i + v_{it}; \\ \varepsilon_i &\sim N\left(0, \frac{\sigma_v^2}{1 - \alpha^2}\right); \eta_i \sim N(0, \sigma_\eta^2); v_{it} \sim N(0, \sigma_v^2), \end{aligned}$$

for sample size  $N = 200$ ;  $\sigma_v^2 = 1$ , and different values of  $\alpha = \{0.4, 0.8\}$  and  $\sigma_\eta^2 = \{\frac{1}{4}, 1, 4\}$ . 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 for  $\alpha = 0.4$  and  $\alpha = 0.8$  respectively.

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 with  $\alpha = 0.8$ . The relative biases of the DIF 2SLS and LEV estimators are, however, the same when  $\sigma_\eta^2 = \sigma_v^2$ . 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  $\sigma_\eta^2$  when  $\alpha = 0.4$ , whereas that of the LEV 2SLS estimator does. It is only 0.029 when  $\sigma_\eta^2 = \frac{1}{4}$ , but increases to 0.169 when  $\sigma_\eta^2 = 4$ . 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  $\sigma_\eta^2 = \frac{1}{4}$  and 13.0 when  $\sigma_\eta^2 = 4$ , 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  $\sigma_\eta^2 = 4$ , 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  $\sigma_\eta^2 = \frac{1}{4}$ , with the expected concentration parameter equal to 20.9. It increases to 0.68 when  $\sigma_\eta^2 = 4$  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  $\alpha$ , becomes more severe with increasing  $\sigma_\eta^2/\sigma_v^2$ . As

both the OLS bias and the relative bias increase with increasing  $\sigma_\eta^2$ , 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  $\sigma_\eta^2 = \frac{1}{4}$  to 0.132 when  $\sigma_\eta^2 = 4$ .

The SYS 2SLS estimator has a slightly smaller relative bias than the DIF and LEV ones when  $\sigma_\eta^2 = \sigma_v^2$ . 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  $\sigma_\eta^2 = \frac{1}{4}$ , 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  $\sigma_\eta^2 = 4$  the relative and absolute biases of the SYS 2SLS estimator are similar to that of the LEV 2SLS estimator, albeit slightly smaller.

Table 1. Cross Section Estimation Results for  $\alpha = 0.4$ ,  $N = 200$ ,  $t = 6$  and  $\sigma_v^2 = 1$

|                               | DIF     |               |         | LEV    |               |         | SYS    |               |         |
|-------------------------------|---------|---------------|---------|--------|---------------|---------|--------|---------------|---------|
|                               | Coeff   | <i>StDev</i>  | RelBias | Coeff  | <i>StDev</i>  | RelBias | Coeff  | <i>StDev</i>  | RelBias |
| $\sigma_\eta^2 = \frac{1}{4}$ |         |               |         |        |               |         |        |               |         |
| OLS                           | -0.3005 | <i>0.0670</i> |         | 0.6208 | <i>0.0555</i> |         | 0.2243 | <i>0.0566</i> |         |
| 2SLS                          | 0.3698  | <i>0.1734</i> | 0.0431  | 0.4064 | <i>0.0915</i> | 0.0289  | 0.3890 | <i>0.0810</i> | 0.0627  |
| $E(\mu)$                      |         | 58.06         |         |        | 132.7         |         |        |               |         |
| $\sigma_\eta^2 = 1$           |         |               |         |        |               |         |        |               |         |
| OLS                           | -0.3005 | <i>0.0670</i> |         | 0.8196 | <i>0.0407</i> |         | 0.5230 | <i>0.0491</i> |         |
| 2SLS                          | 0.3637  | <i>0.1892</i> | 0.0518  | 0.4240 | <i>0.1131</i> | 0.0572  | 0.4038 | <i>0.0953</i> | 0.0306  |
| $E(\mu)$                      |         | 46.75         |         |        | 46.75         |         |        |               |         |
| $\sigma_\eta^2 = 4$           |         |               |         |        |               |         |        |               |         |
| OLS                           | -0.3005 | <i>0.0670</i> |         | 0.9416 | <i>0.0239</i> |         | 0.8118 | <i>0.0292</i> |         |
| 2SLS                          | 0.3604  | <i>0.1973</i> | 0.0566  | 0.4917 | <i>0.1565</i> | 0.1694  | 0.4622 | <i>0.1223</i> | 0.1511  |
| $E(\mu)$                      |         | 42.31         |         |        | 13.02         |         |        |               |         |

Table 2. Cross Section Estimation Results for  $\alpha = 0.8$ ,  $N = 200$ ,  $t = 6$  and  $\sigma_v^2 = 1$

|                               | DIF     |               |         | LEV    |               |         | SYS    |               |         |
|-------------------------------|---------|---------------|---------|--------|---------------|---------|--------|---------------|---------|
|                               | Coeff   | <i>StDev</i>  | RelBias | Coeff  | <i>StDev</i>  | RelBias | Coeff  | <i>StDev</i>  | RelBias |
| $\sigma_\eta^2 = \frac{1}{4}$ |         |               |         |        |               |         |        |               |         |
| OLS                           | -0.1003 | <i>0.0699</i> |         | 0.9382 | <i>0.0246</i> |         | 0.8239 | <i>0.0281</i> |         |
| 2SLS                          | 0.5973  | <i>0.4041</i> | 0.2251  | 0.8150 | <i>0.0841</i> | 0.1088  | 0.7925 | <i>0.0825</i> | 0.3136  |
| $E(\mu)$                      |         | 9.15          |         |        | 20.92         |         |        |               |         |
| $\sigma_\eta^2 = 1$           |         |               |         |        |               |         |        |               |         |
| OLS                           | -0.1003 | <i>0.0699</i> |         | 0.9798 | <i>0.0142</i> |         | 0.9380 | <i>0.0153</i> |         |
| 2SLS                          | 0.5210  | <i>0.4636</i> | 0.3100  | 0.8562 | <i>0.0920</i> | 0.3123  | 0.8336 | <i>0.0901</i> | 0.2433  |
| $E(\mu)$                      |         | 6.35          |         |        | 6.35          |         |        |               |         |
| $\sigma_\eta^2 = 4$           |         |               |         |        |               |         |        |               |         |
| OLS                           | -0.1003 | <i>0.0699</i> |         | 0.9945 | <i>0.0074</i> |         | 0.9827 | <i>0.0074</i> |         |
| 2SLS                          | 0.4844  | <i>0.4852</i> | 0.3505  | 0.9324 | <i>0.0852</i> | 0.6808  | 0.9169 | <i>0.0785</i> | 0.6396  |
| $E(\mu)$                      |         | 5.45          |         |        | 1.68          |         |        |               |         |

Figures 3 and 4 display p-value plots for the Wald test for testing  $H_0 : \alpha = \alpha_0$  with  $\alpha_0$  the true parameter value. When  $\sigma_\eta^2 = \sigma_v^2 = 1$ , the size properties of the Wald tests based on the DIF and LEV 2SLS estimates are virtually identical, which is as expected as the concentration parameters are equal in expectation. 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  $\sigma_\eta^2 = \frac{1}{4}$  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  $\sigma_\eta^2 = 4$ . The Wald test results based on the DIF 2SLS estimates are not very sensitive to the value of  $\sigma_\eta^2$ . These results are again in line with expectation given the results of the previous section.

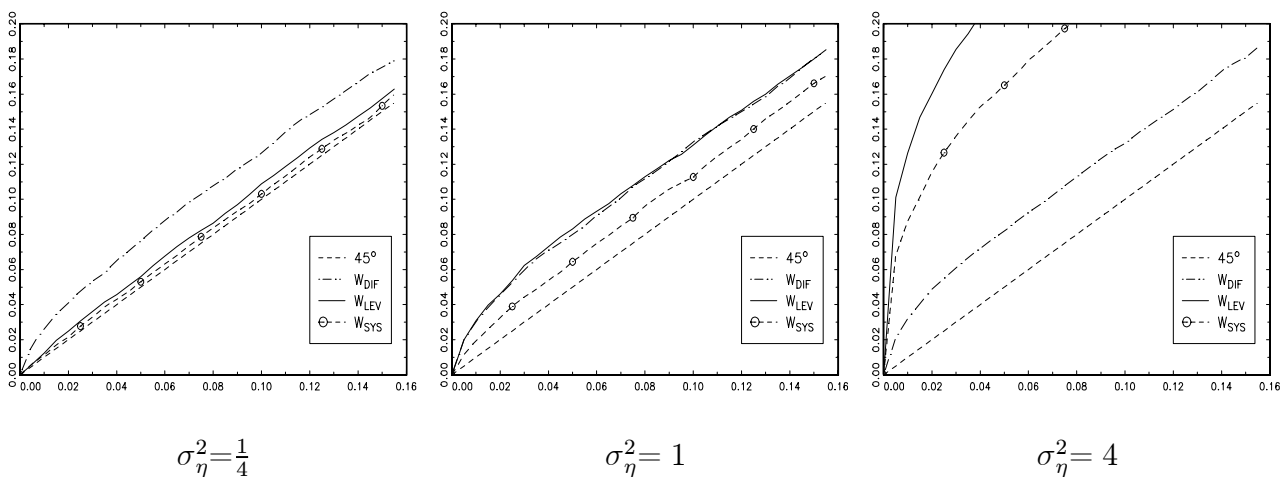


Figure 3. P-value plots, Wald test  $H_0 : \alpha = 0.4$ .

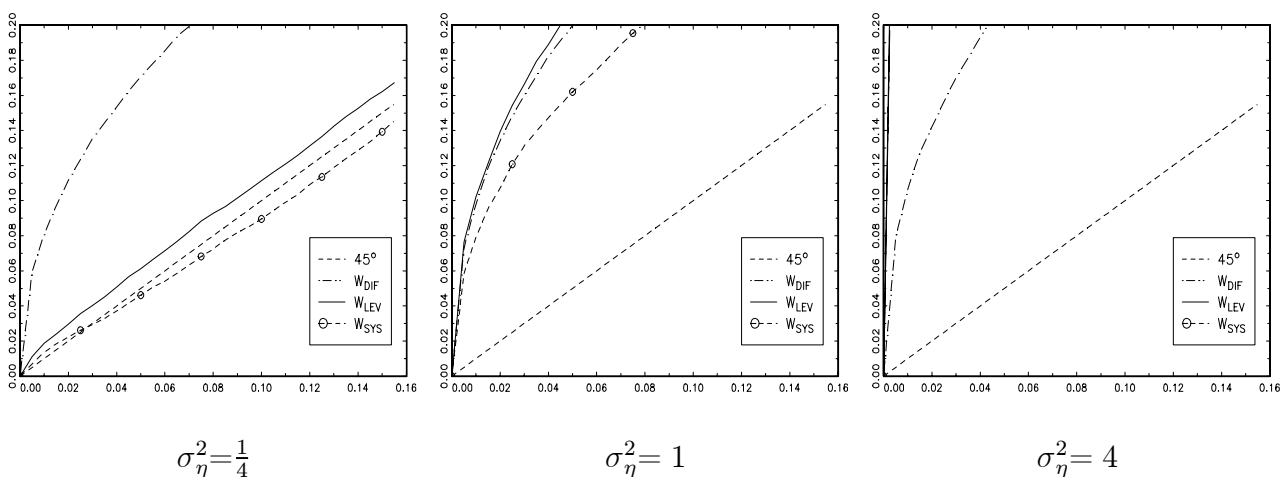


Figure 4. P-value plots, Wald test  $H_0 : \alpha = 0.8$ .

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

$$E\left(\frac{1}{N}\mu_{l3}\right) > E\left(\frac{1}{N}\mu_{d3}\right) \text{ if } \sigma_\varepsilon^2 < \frac{\sigma_v^2}{1 - \alpha^2}$$

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

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.

Table 3 presents Monte Carlo estimation results for the AR(1) model with normally distributed  $\eta_i$  and  $v_i$ , with  $N = 200$ ,  $T = 6$ ,  $\alpha = 0.8$  and  $\sigma_v^2 = 1$ , varying  $\sigma_\eta^2 = \{\frac{1}{4}, 1, 4\}$ . We 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,  $-1$ s 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 2. 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  $\sigma_\eta^2 = 1$  and  $\sigma_\eta^2 = 4$ , 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.

Table 3. Panel Data Estimation Results,  $N = 200$ ,  $T = 6$ ,  $\alpha = 0.8$  and  $\sigma_v^2 = 1$

|                               | DIF     |        |         | LEV    |        |         | SYS    |        |         |
|-------------------------------|---------|--------|---------|--------|--------|---------|--------|--------|---------|
|                               | Coeff   | StDev  | RelBias | Coeff  | StDev  | RelBias | Coeff  | StDev  | RelBias |
| $\sigma_\eta^2 = \frac{1}{4}$ |         |        |         |        |        |         |        |        |         |
| OLS                           | -0.0999 | 0.0327 |         | 0.9382 | 0.0114 |         | 0.8238 | 0.0182 |         |
| 2SLS                          | 0.5807  | 0.1624 | 0.2437  | 0.8119 | 0.0561 | 0.0858  | 0.7789 | 0.0736 | 0.8866  |
| 1-step                        | 0.7338  | 0.1306 |         |        |        |         | 0.7983 | 0.0672 |         |
| 2-step                        | 0.7336  | 0.1403 |         | 0.8117 | 0.0598 |         | 0.7973 | 0.0596 |         |
| $\sigma_\eta^2 = 1$           |         |        |         |        |        |         |        |        |         |
| OLS                           | -0.0999 | 0.0327 |         | 0.9799 | 0.0063 |         | 0.9381 | 0.0086 |         |
| 2SLS                          | 0.4692  | 0.2122 | 0.3675  | 0.8502 | 0.0679 | 0.2792  | 0.8129 | 0.0792 | 0.0932  |
| 1-step                        | 0.6721  | 0.1814 |         |        |        |         | 0.8299 | 0.0730 |         |
| 2-step                        | 0.6639  | 0.2009 |         | 0.8438 | 0.0424 |         | 0.8182 | 0.0684 |         |
| $\sigma_\eta^2 = 4$           |         |        |         |        |        |         |        |        |         |
| OLS                           | -0.0999 | 0.0327 |         | 0.9946 | 0.0032 |         | 0.9828 | 0.0036 |         |
| 2SLS                          | 0.4012  | 0.2395 | 0.4431  | 0.9239 | 0.0635 | 0.6369  | 0.8891 | 0.0751 | 0.4875  |
| 1-step                        | 0.6175  | 0.2131 |         |        |        |         | 0.8997 | 0.0700 |         |
| 2-step                        | 0.6007  | 0.2410 |         | 0.9133 | 0.0792 |         | 0.8841 | 0.0789 |         |

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  $\sigma_\eta^2 = \frac{1}{4}$ , especially for the 1-step SYS GMM estimator. The size behaviours are very similar when  $\sigma_\eta^2 = 1$ , but the SYS Wald tests size properties are much worse than that of the DIF Wald tests when  $\sigma_\eta^2 = 4$ .

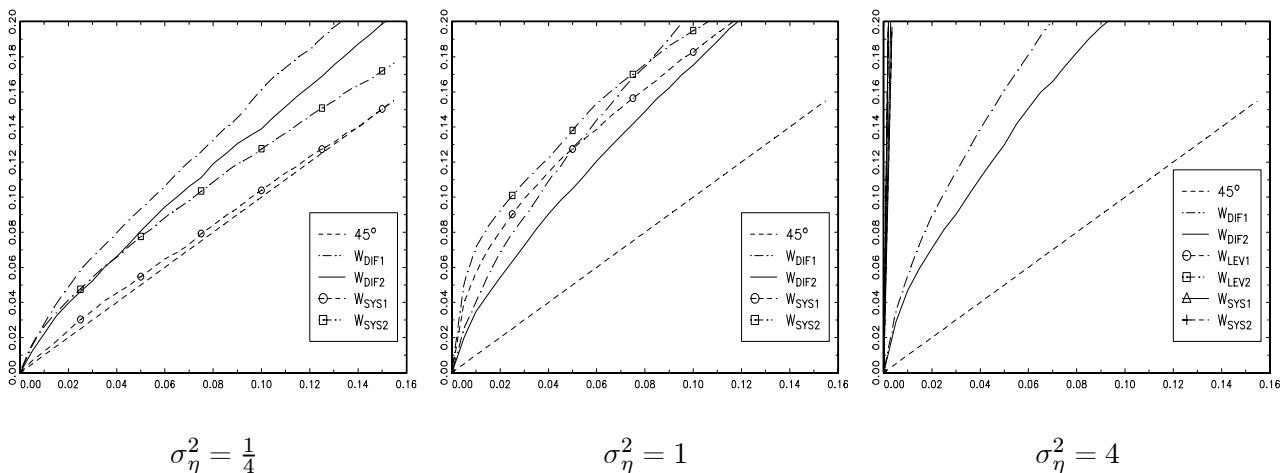


Figure 5. P-value plots, Wald test  $H_0 : \alpha = 0.8$ .

## 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 ( $\sigma_\eta^2$ ) and idiosyncratic errors ( $\sigma_v^2$ ) 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  $\sigma_\eta^2 \geq \sigma_v^2$ , 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  $\sigma_\eta^2 = \sigma_v^2$ . We show 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 (2007) 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 (2007) 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

$$y_{it} = x_{it}\beta + \eta_i + v_{it}$$

$$x_{it} = \rho x_{i,t-1} + \gamma \eta_i + \delta v_{it} + w_{it}$$

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.

## References

- [1] 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.
- [2] Arellano, M. and O. Bover (1995), Another Look at the Instrumental Variable Estimation of Error-Components Models, *Journal of Econometrics* 68, 29-51.
- [3] Blundell, R. and S. Bond, 1998, Initial Conditions and Moment Restrictions in Dynamic Panel Data Models, *Journal of Econometrics* 87, 115-143.
- [4] 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, JAI Press, Elsevier Science.
- [5] Bond, S.R., A. Hoeffler and J. Temple (2001). GMM Estimation of Empirical Growth Models, Working Paper, University of Oxford.

- [6] Bond, S.R. and F. Windmeijer (2005), Reliable Inference for GMM Estimators? Finite Sample Properties of Alternative Test Procedures in Linear Panel Data Models, *Econometrics Reviews* 24, 1-37.
- [7] 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.
- [8] 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.
- [9] 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.
- [10] Han, C. and P.C.B. Phillips (2006), GMM with Many Moment Conditions, *Econometrica* 74, 147-192.
- [11] 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.
- [12] Hayakawa, K. (2005), Small Sample Bias Properties of the System GMM Estimator in Dynamic Panel Data Models, Hi-Stat Discussion Paper Series d05-82, Institute of Economic Research, Hitotsubashi University.
- [13] Kleibergen, F. (2005) Testing Parameters in GMM Without Assuming They are Identified, *Econometrica* 73, 1103-1123.

- [14] Levine, R., L. Norman and T. Beck (2000), Financial Intermediation and Growth: Causality and Causes, *Journal of Monetary Economics* 46, 31-77.
- [15] Levinsohn, J. and A. Petrin (2003), Estimating Production Functions Using Inputs to Control for Unobservables, *Review of Economic Studies* 70, 317-341.
- [16] Moreira, M. (2003), A Conditional Likelihood Ratio Test for Structural Models, *Econometrica* 71, 1027-1048.
- [17] Newey, W.K. and F. Windmeijer (2007), GMM with Many Weak Moment Conditions, working paper, MIT.
- [18] 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.
- [19] Ridder, G. and T. Wansbeek (1990), Dynamic Models for Panel Data, in F. van der Ploeg (ed.), *Advanced Lectures in Quantitative Economics*, Academic Press.
- [20] Rothenberg, T.J. (1984), Approximating the Distributions of Econometric Estimators and Test Statistics, in Z. Griliches and M.D. Intriligator (eds.), *Handbook of Econometrics*, Vol. II, North Holland, 881-935.
- [21] Staiger, D. and J.H. Stock (1997), Instrumental Variables Regression with Weak Instruments, *Econometrica* 65, 557-586.
- [22] Stock, J.H. and J.H. Wright (2000), GMM with Weak Identification, *Econometrica* 68, 1055-1096.
- [23] 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*, 518-529.

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

## 7 Appendix

### 7.1 Concentration parameters in cross-section analysis

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

$$\begin{aligned}\Delta y_{it} &= \alpha \Delta y_{i,t-1} + \Delta u_{it} \\ \Delta y_{i,t-1} &= y_i^{t-2'} \pi_{dt} + d_{i,t-1}^t.\end{aligned}$$

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

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

but as

$$\pi_{dt} = [E(y_i^{t-2} y_i^{t-2'})]^{-1} E(y_i^{t-2} \Delta y_{i,t-1})$$

and

$$\sigma_{dt}^2 = E\left((\Delta y_{i,t-1} - y_i^{t-2'} \pi_{dt})^2\right)$$

we get

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

Under covariance stationarity

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

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  $E(y_i^{t-2}y_i^{t-2'})$  is given by (see e.g. Ridder and Wansbeek (1990))

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

where

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

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

We further have that

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

where

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

As

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

and so

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

Further

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

Combining these results in

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

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

$$\begin{aligned} y_{it} &= \alpha y_{i,t-1} + \eta_i + v_{it} \\ y_{i,t-1} &= \Delta y_i^{t-1'} \pi_{it} + l_{i,t-1}^t \end{aligned}$$

and the expected concentration parameter is given by

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

Again, under covariance stationarity, we have that

$$E(\Delta y_i^{t-1}\Delta y_i^{t-1'}) = \frac{\sigma_v^2}{1+\alpha} \begin{bmatrix} 2 & \alpha-1 & \alpha(\alpha-1) & \cdots & \alpha^{t-4}(\alpha-1) \\ \alpha-1 & 2 & \alpha-1 & \cdots & \alpha^{t-5}(\alpha-1) \\ \alpha(\alpha-1) & \alpha-1 & 2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \alpha-1 \\ \alpha^{t-4}(\alpha-1) & \cdots & \alpha(\alpha-1) & \alpha-1 & 2 \end{bmatrix}$$

and

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

It then follows that

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

As

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

we get that

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

## 7.2 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  $E[\varepsilon_i^2] = \sigma_\varepsilon^2$ .

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

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

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

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

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

For the levels model we get

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

and

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

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

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

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.