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

Maurice Bun
Tinbergen Institute, University of Amsterdam

Frank Windmeijer
University of Bristol, CMPO & Cemmap
• Arellano-Bond GMM (DIF) estimator suffers weak instrument bias when series are persistent.

• Blundell-Bond proposed “System” (SYS) GMM estimator that is less biased, combining differences with levels information.

• The concentration parameters in the cross sections can be shown to be the same for the differenced and levels moment conditions under covariance stationarity and certain parameter configurations.

• Relative bias and Wald test performances are then similar for DIF and LEV 2SLS estimators in the cross section.

• Results carry over to SYS estimator and for analysis on the whole panel, estimating by GMM.
Dynamic Panel Data Model

\[ y_{it} = \alpha y_{i,t-1} + u_{it} \]
\[ u_{it} = \eta_i + v_{it} \]

\[ i = 1, \ldots, N; t = 2, \ldots, T \]

estimation for finite T and large N:
1. pooled OLS upward biased
2. within groups downward biased
3. first differenced OLS downward biased
4. IV/GMM/ML consistent
GMM

Arellano and Bond (1991):

first-difference the model (DIF) and use lagged levels as instruments

Blundell and Bond (1998):

additional moment conditions for model in levels (LEV) with lagged differences as instruments when initial observations obey mean stationarity
example for T=3:

\[ \Delta y_{i3} = \alpha \Delta y_{i,2} + \Delta v_{i3} \]

\[ \Delta y_{i,2} = y_{i1} \pi_d + d_i \]

\[ y_{i3} = \alpha y_{i,2} + \eta_i + v_{i3} \]

\[ y_{i,2} = \Delta y_{i2} \pi_l + l_i \]
Blundell and Bond (1998):

1. $\hat{\pi}_d \to 0$ when $\alpha \to 1$ and/or $\sigma^2_\eta / \sigma^2_v \to \infty$

2. $\hat{\pi}_l \to 0.5$

3. it seems that there is no weak instrument problem in the LEV model and simulations show that the LEV (and SYS) estimator has smaller bias and rmse than the DIF estimator

<table>
<thead>
<tr>
<th>$\alpha = 0.8, N = 200, T = 6$</th>
<th>DIF</th>
<th>SYS</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\sigma^2_\eta = \sigma^2_v$</td>
<td>coeff</td>
<td>St dev</td>
</tr>
<tr>
<td>2-step GMM</td>
<td>0.66</td>
<td>0.20</td>
</tr>
</tbody>
</table>
concentration parameter

simple linear IV model

\[ y_i = \beta x_i + u_i \]
\[ x_i = z_i' \pi + \xi_i, \]

where \( x_i \) is endogenous regressor, \( z_i \) are \( k_z \) instruments and \( \beta \) is estimated by 2SLS.

concentration parameter is measure of strength of the instruments defined as

\[ \mu = \frac{\pi' Z' Z \pi}{\sigma_{\xi}^2} \]
concentration parameter evaluated at first-stage OLS estimates $\hat{\pi}$ and $\hat{\sigma}_\xi^2$, and divided by the number of instruments $k_z$ is the first-stage F-statistic for testing whether $\pi = 0$.

when concentration parameter is small we have weak instruments, leading to bias of $\hat{\beta}_{2SLS}$ towards OLS and poor size properties of the Wald test.
2SLS bias approximation:

\[ E\left[ \hat{\beta}_{2SLS} - \beta \right] \approx \frac{\sigma_{u\xi}}{\sigma_{\xi}^2} \frac{k_z - 2}{\mu} \]

Relative bias of 2SLS to OLS depends on \( k_z \) and \( \mu \) only

Size distortion Wald test depends on \( \rho_{u\xi} \) and \( \mu \)
application to dynamic panel data model (for cross-section at time $t$):

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

$$\Delta y_{i,t-1} = y_{i}^{t-2}' \pi_{dt} + d_{i,t-1}^t$$

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

$$y_{i,t-1} = \Delta y_{i}^{t-1}' \pi_{lt} + l_{i,t-1}^t$$
expected values of concentration parameters in DIF and LEV models for the cross-section at time $t$ are

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

$$E\left(\frac{1}{N} \mu_{lt}\right) = \frac{\pi_l'E \left( \Delta y_{i}^{t-1} \Delta y_{i}^{t-1'} \right) \pi_{lt}}{\sigma_{lt}^2}$$
in the paper we prove that (covariance stationary initial condition)

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

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

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

\[
E\left(\frac{1}{N} \mu_{lt}\right) = \frac{1}{t-2} \left( 1 + (t-3)\frac{\sigma_\eta^2}{\sigma_v^2} \right)
\]
it then follows that

\[ E\left( \frac{1}{N} \mu_{dt} \right) = E\left( \frac{1}{N} \mu_{lt} \right) \quad \text{if} \quad \sigma^2_\eta = \sigma^2_v \]

\[ E\left( \frac{1}{N} \mu_{dt} \right) \geq E\left( \frac{1}{N} \mu_{lt} \right) \quad \text{if} \quad \sigma^2_\eta \geq \sigma^2_v, \quad t > 3 \]
questions:

Why does the system estimator perform so much better if the concentration parameters for the DIF and LEV moment conditions are the same when $\sigma^2_{\eta} = \sigma^2_v$?

How does the weak instrument problem manifest itself in the LEV and SYS models/estimators?
Further cross-section analysis

DIF and LEV models have very different endogeneity features.

DIF:

$$\Delta y_t = \alpha \Delta y_{t-1} + \Delta v_t$$

LEV:

$$y_t = \alpha y_{t-1} + \eta + v_t$$

Estimated by OLS we get much larger bias for OLS DIF (negative), than for OLS LEV (positive) for $\alpha$ large.
SYS “model” can be written as

\[
\begin{pmatrix}
\Delta y_t \\
y_t
\end{pmatrix} = \alpha \begin{pmatrix}
\Delta y_{t-1} \\
y_{t-1}
\end{pmatrix} + \begin{pmatrix}
\Delta \eta \\
\eta + v_t
\end{pmatrix}
\]

OLS on this model results in weighted average of DIF and LEV OLS estimators:

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

\[
\tilde{\gamma} = \frac{\Delta y'_{t-1}\Delta y_{t-1}}{\Delta y'_{t-1}\Delta y_{t-1} + y'_{t-1}y_{t-1}} ; \quad \text{plim } \tilde{\gamma} = \frac{1 - \alpha}{\frac{3}{2} - \alpha + \frac{1}{2} \frac{\sigma^2}{\sigma^2} \frac{1 + \alpha}{1 - \alpha}}
\]
asymptotic bias of OLS
OLS bias for LEV/SYS is much smaller than that of DIF. Therefore LEV/SYS IV estimator will have less bias than DIF IV, even if instruments are completely non-informative.

Stock and Yogo (2005) show, however, that we expect relative biases and size distortions of Wald tests to be similar for DIF and LEV 2SLS in the cross-section when the concentration parameters are the same.

remark: SYS 2SLS estimator is weighted average of the DIF and LEV 2SLS estimators.
Some Monte Carlo results for cross-section $t=6$, $\alpha = 0.4$, $N = 200$

<table>
<thead>
<tr>
<th>$\sigma^2_n$</th>
<th>DIF</th>
<th>LEV</th>
<th>SYS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Coeff StDev RelB</td>
<td>Coeff StDev RelB</td>
<td>Coeff StDev RelB</td>
<td></td>
</tr>
<tr>
<td><strong>$\sigma^2_n = 0.25$</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>OLS</td>
<td>-0.301 0.067</td>
<td>0.621 0.056</td>
<td>0.224 0.057</td>
</tr>
<tr>
<td>2SLS</td>
<td>0.370 0.173 <strong>0.043</strong></td>
<td>0.406 0.092 <strong>0.029</strong></td>
<td>0.389 0.081 <strong>0.063</strong></td>
</tr>
<tr>
<td>$E(\mu)$</td>
<td>58.06</td>
<td>132.7</td>
<td></td>
</tr>
<tr>
<td><strong>$\sigma^2_n = 1$</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>OLS</td>
<td>-0.301 0.067</td>
<td>0.820 0.041</td>
<td>0.523 0.049</td>
</tr>
<tr>
<td>2SLS</td>
<td>0.364 0.189 <strong>0.052</strong></td>
<td>0.424 0.113 <strong>0.057</strong></td>
<td>0.404 0.095 <strong>0.031</strong></td>
</tr>
<tr>
<td>$E(\mu)$</td>
<td>46.75</td>
<td>46.75</td>
<td></td>
</tr>
<tr>
<td><strong>$\sigma^2_n = 4$</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>OLS</td>
<td>-0.301 0.067</td>
<td>0.942 0.024</td>
<td>0.812 0.029</td>
</tr>
<tr>
<td>2SLS</td>
<td>0.360 0.197 <strong>0.057</strong></td>
<td>0.492 0.157 0.169</td>
<td>0.462 0.122 <strong>0.151</strong></td>
</tr>
<tr>
<td>$E(\mu)$</td>
<td>42.31</td>
<td>13.02</td>
<td></td>
</tr>
</tbody>
</table>
Some Monte Carlo results for cross-section $t=6$, $\alpha = 0.8$, $N = 200$

<table>
<thead>
<tr>
<th></th>
<th>DIF</th>
<th></th>
<th>LEV</th>
<th></th>
<th>SYS</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Coeff</td>
<td>StDev</td>
<td>RelB</td>
<td>Coeff</td>
<td>StDev</td>
</tr>
<tr>
<td>$\sigma^2_\eta = 0.25$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>OLS</td>
<td>-0.100</td>
<td>0.070</td>
<td></td>
<td>0.938</td>
<td>0.025</td>
</tr>
<tr>
<td>2SLS</td>
<td>0.597</td>
<td>0.404</td>
<td>0.225</td>
<td>0.815</td>
<td>0.084</td>
</tr>
<tr>
<td>$E(\mu)$</td>
<td>9.15</td>
<td></td>
<td></td>
<td>20.92</td>
<td></td>
</tr>
<tr>
<td>$\sigma^2_\eta = 1$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>OLS</td>
<td>-0.100</td>
<td>0.070</td>
<td></td>
<td>0.980</td>
<td>0.014</td>
</tr>
<tr>
<td>2SLS</td>
<td>0.521</td>
<td>0.464</td>
<td>0.310</td>
<td>0.856</td>
<td>0.092</td>
</tr>
<tr>
<td>$E(\mu)$</td>
<td>6.35</td>
<td></td>
<td></td>
<td>6.35</td>
<td></td>
</tr>
<tr>
<td>$\sigma^2_\eta = 4$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>OLS</td>
<td>-0.100</td>
<td>0.070</td>
<td></td>
<td>0.995</td>
<td>0.007</td>
</tr>
<tr>
<td>2SLS</td>
<td>0.484</td>
<td>0.485</td>
<td>0.351</td>
<td>0.932</td>
<td>0.085</td>
</tr>
<tr>
<td>$E(\mu)$</td>
<td>5.45</td>
<td></td>
<td></td>
<td>1.68</td>
<td></td>
</tr>
</tbody>
</table>
Panel Data Analysis

no general extension of concept of concentration parameter to panel data framework estimated by efficient GMM

2SLS on the panel is weighted average of period specific 2SLS estimates

bias depends on the weight matrix used: 2SLS goes to OLS, but GMM using the efficient weight matrix under homoskedasticity for DIF, taking account of the MA(1) errors, goes to within groups

preliminary results show that cross-sectional relative bias results extend to panel 2SLS when both T and N are large
Some Monte Carlo results for Panel $\sigma_{\eta}^2 = 1$, $\alpha = 0.8$, $N = 200$

<table>
<thead>
<tr>
<th></th>
<th>DIF</th>
<th>LEV</th>
<th>SYS</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Coeff</td>
<td>StDev</td>
<td>RelB</td>
</tr>
<tr>
<td>$T = 6$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>OLS</td>
<td>-0.100</td>
<td>0.033</td>
<td></td>
</tr>
<tr>
<td>2SLS</td>
<td>0.469</td>
<td>0.212</td>
<td>0.367</td>
</tr>
<tr>
<td>1-step</td>
<td>0.672</td>
<td>0.181</td>
<td></td>
</tr>
<tr>
<td>2-step</td>
<td>0.664</td>
<td>0.200</td>
<td></td>
</tr>
<tr>
<td>$T = 15$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>OLS</td>
<td>-0.100</td>
<td>0.019</td>
<td></td>
</tr>
<tr>
<td>2SLS</td>
<td>0.374</td>
<td>0.075</td>
<td>0.473</td>
</tr>
<tr>
<td>1-step</td>
<td>0.757</td>
<td>0.040</td>
<td></td>
</tr>
<tr>
<td>2-step</td>
<td>0.754</td>
<td>0.046</td>
<td></td>
</tr>
</tbody>
</table>
25
Conclusions

• Theoretical results show that cross-section based concentration parameters are same for DIF and LEV when variance ratio is 1.

• Simulation results show that system GMM estimator also suffers from a weak instrument problem in the panel AR(1) model with persistent data.

• This does not establish itself in large bias (and variance) as for DIF, but in the relative bias and Wald test performance.

• Important to make $\sigma_{\eta}^2$ small, i.e. add individual specific variables to the model in levels when estimating by SYS.