Dynamics and Measurement Error in Panel Data Signal-Noise Pattern, Heterogeneity and GMM: A Case-Study of the FDI Impact on GDP Growth

Erik Biorn, University of Oslo (Coauthor)

Xuehui Han, Asian Development Bank (Presenter)

20th International Conference on Panel Data,
Hitotsubashi, Tokyo, July 9–10, 2014

Introduction

- Sources of biases in OLS estimation:
  - (A) Mis-measured regressors: EIV. 'Error'=Measurement error
  - (B) Lagged endogenous regressors in conjunction with autocorrelated disturbances – ARMA
- *THIS PAPER:* Finite sample properties of GMM on PD, when (A) & (B) occur: *Monte Carlo Simulations*.
- PRESENTATION:
  - 1. Model framework
  - 2. Simulation model design
  - 3. Results
  - 4. FDI Case

Why be interested in memory in errors in panel data, not only in memory in disturbances?

Examples:

- Stock variable obtained by cumulating flows – Perpetual Inventory Method:

  $\Longrightarrow$ *Measurement errors varying cyclically.*

- Flow variable: Improper periodization of transactions.

  $\Longrightarrow$ *Serial correlation between errors which are close in time.*

EIV-ARMAX-PD model:

$$
\begin{aligned}
\mu_{it} &= \alpha_i + \boldsymbol{\xi}_{it}\boldsymbol{\beta} + \mu_{i,t-1}\lambda + u_{it}, && |\lambda| < 1,\\
y_{it} &= \mu_{it} + \nu_{it},\\
\mathbf{q}_{it} &= \boldsymbol{\xi}_{it} + \boldsymbol{\eta}_{it},\\
\boldsymbol{\xi}_{it} &\perp u_{it} \perp \nu_{it} \perp \boldsymbol{\eta}_{it}, && i = 1, \dots, N;\ t = 1, \dots, T.
\end{aligned}
$$

Latent: $\mu_{it}, \boldsymbol{\xi}_{it}$ (vector). Observed: $y_{it}, \mathbf{q}_{it}$ (vector).

α_i = fixed effect (or random, correlated with $\boldsymbol{\xi}_{it}$)

Memory, represented by V(MA) in simulations.

$$
\begin{aligned}
&u_{it} : \text{ zero or finite memory } N_u,\\
&\nu_{it} : \text{ zero or finite memory } N_\nu,\\
&\boldsymbol{\eta}_{it} : \text{ zero or finite memory } N_\eta,\\
&\boldsymbol{\xi}_{it} = \chi_i \text{ (time invariant) } + \psi_{it} \text{ (finite memory } N_\xi),
\end{aligned}
$$

Equations after elimination of latent variables:

LEVEL FORM:

$$y_{it} = \alpha_i + \mathbf{q}_{it}\boldsymbol{\beta} + y_{i,t-1}\lambda + w_{it},$$

$$w_{it} = u_{it} + \nu_{it} - \nu_{i,t-1}\lambda - \boldsymbol{\eta}_{it}\boldsymbol{\beta} = \text{Composite error.}$$

DIFFERENCE FORM:

$$\Delta y_{it} = \Delta\mathbf{q}_{it}\boldsymbol{\beta} + \Delta y_{i,t-1}\lambda + \Delta w_{it},$$

$$\Delta w_{it} = \Delta u_{it} + \Delta\nu_{it} - \Delta\nu_{i,t-1}\lambda - \Delta\boldsymbol{\eta}_{it}\boldsymbol{\beta} = \text{Composite error.}$$

Reminder: Differencing eliminates time invariant variables: $\boldsymbol{\alpha}_i$ & χ_i

IV SET FOR $\mathbf{x}_{it} = (\mathbf{q}_{it}, y_{i,t-1})$ & $\Delta\mathbf{x}_{it} = (\Delta\mathbf{q}_{it}, \Delta y_{i,t-1})$

MUST SATISFY

- *Orthogonality condition*: IVs orthogonal to Composite disturbance $(w_{it}, \Delta w_{it})$.
- *Rank condition*: IVs correlated with Instrumented variables.

AND TAKE ACCOUNT OF

- Memory of Observed regressor $\mathbf{q}_{it} = \max[N_\xi, N_\eta]$.
- Memory of Gross disturbance $w_{it} = \max[N_u, N_\nu + 1, N_\eta]$.

Let $\gamma = [\beta', \lambda]'$, $\mathbf{x}_{it} = (\mathbf{q}_{it}, y_{i,t-1})$ and write the respective equations as

$$y_{it} = \alpha_i + \mathbf{x}_{it}\gamma + w_{it},$$
$$\Delta y_{it} = \Delta \mathbf{x}_{it}\gamma + \Delta w_{it},$$

Let L =Level, D =Difference and stack the equations into:

$$\mathbf{y}_{Li} = \alpha_i + \mathbf{X}_{Li}\gamma + \mathbf{w}_{Li}, \qquad \mathbf{Z}_{Di} = \mathbf{Q}\mathbf{X}_{Di} \longrightarrow \text{ IV for } \mathbf{X}_{Li},$$
$$\mathbf{y}_{Di} = \mathbf{X}_{Di}\gamma + \mathbf{w}_{Di}, \qquad \mathbf{Z}_{Li} = \mathbf{P}\mathbf{X}_{Li} \longrightarrow \text{ IV for } \mathbf{X}_{Di}.$$

$\mathbf{Q}, \mathbf{P}$ = selection matrices for respective IV sets.

Boldface: Vectors/Matrices.

Two-step GMM: Let $\widehat{\mathbf{w}}_{Li}$ & $\widehat{\mathbf{w}}_{Di}$ = residuals from 'step-one' GMM.

GMM for Eq. in LEVELS:

$$\widetilde{\gamma}_L = \{[\textstyle\sum_i \mathbf{X}'_{Li}\mathbf{Z}_{Di}][\sum_i \mathbf{Z}'_{Di}\widehat{\mathbf{w}}_{Li}\widehat{\mathbf{w}}'_{Li}\mathbf{Z}_{Di}]^{-1}[\sum_i \mathbf{Z}'_{Di}\mathbf{X}_{Li}]\}^{-1} \times \{[\sum_i \mathbf{X}'_{Li}\mathbf{Z}_{Di}][\sum_i \mathbf{Z}'_{Di}\widehat{\mathbf{w}}_{Li}\widehat{\mathbf{w}}'_{Li}\mathbf{Z}_{Di}]^{-1}[\sum_i \mathbf{Z}'_{Di}\mathbf{y}_{Li}]\},$$

GMM for Eq. in DIFFERENCES:

$$\widetilde{\gamma}_D = \{[\textstyle\sum_i \mathbf{X}'_{Di}\mathbf{Z}_{Li}][\sum_i \mathbf{Z}'_{Li}\widehat{\mathbf{w}}_{Di}\widehat{\mathbf{w}}'_{Di}\mathbf{Z}_{Li}]^{-1}[\sum_i \mathbf{Z}'_{Li}\mathbf{X}_{Di}]\}^{-1} \times \{[\sum_i \mathbf{X}'_{Di}\mathbf{Z}_{Li}][\sum_i \mathbf{Z}'_{Li}\widehat{\mathbf{w}}_{Di}\widehat{\mathbf{w}}'_{Di}\mathbf{Z}_{Li}]^{-1}[\sum_i \mathbf{Z}'_{Li}\mathbf{y}_{Di}]\}.$$

Note: Symmetry between L & D subscripts.

Issues:

- Contrast GMM's performance in AR-EIV models and in boundary cases with a strict AR model and with a static EIV model without autoregression;
- The impact of changed noise and signal pattern on short-run versus long-run coefficient estimates;
- The impact on estimator bias of increased time-invariant heterogeneity in the equation and in the exogenous signal;
- The impact on estimator bias of changed panel size and time-series length.

Simulation Design

The memory pattern is represented by Moving Average:
The signal vector ξ_{it} is a time invariant vector plus a VMA process;
The measurement errors in exogenous and endogenous variables and disturbances are all generated by moving average process;

$(\beta_1, \beta_2, \lambda)$=(0.6,0.3,0.8);

$(\bar{\chi}_1, \bar{\chi}_2) = (5{,}10)$;

${\sigma_\psi}^2$=1;

${\sigma_\chi}^2$=0.1;

${\sigma_\epsilon}^2$=${\sigma_\nu}^2$=${\sigma_u}^2$=${\sigma_\delta}^2$=0.1;

${\sigma_\alpha}^2$=0.1;

Simulation Results

- Issue I:

AR without measurement error:

- Negative bias for βs and positive bias for λ for equations in level;
- Negative bias for βs and λ for equations in difference;
- Smaller bias for case with larger signal spread.

Benchmark Model, Static $\lambda = 0$

- For both equations, βs are negatively biased;
- Introducing error memory or reducing the signal variance make the negative biases larger;
- Positive bias for λ for equations in level and negative bias for λ for equations in difference.

Issue I Cont.: Interacting dynamics with measurement errors

- With $\lambda = 0.8$, for the equation in levels, the effect of introducing measurement error in the exogenous variables only $(\sigma_\epsilon{}^2, \sigma_\delta{}^2) = (0.1, 0.0)$ is an increased bias for both βs and λ, while introduction of measurement errors in the endogenous variable $(\sigma_\epsilon{}^2, \sigma_\delta{}^2) = (0.0, 0.1)$ reduces the bias;
- For the equations in differences, $(\sigma_\epsilon{}^2, \sigma_\delta{}^2) = (0.1, 0.0)$ increase bias, but now $(\sigma_\epsilon{}^2, \sigma_\delta{}^2) = (0.0, 0.1)$magnifies the bias too.

Issue 2: Impact on long-run coefficient

- Compared with the short term coefficients, the long-run coefficients $\frac{\beta_1}{1-\lambda}$ and $\frac{\beta_2}{1-\lambda}$:
  - The bias does not become smaller under large signal spread than under small signal spread ${\sigma_\psi}^2$ invariably as in the short term coef. case;
  - The constant negative bias does not hold either;
  - The negative bias in the long-run coef. for equation in difference still exist.
- When the long-run impacts are parameters of crucial interests in analyzing genuine data, the better choice seems to be equation in levels.

Issue 3: Increased time-invariant heterogeneity

- For equations in levels, an increased equation heterogeneity, ${\sigma_\alpha}^2$, has ambiguous effects on βs and λ, but the negative bias of βs and positive bias of λ remain;
- For equations in levels, an increased signal heterogeneity, ${\sigma_\chi}^2$, leads to larger negative bias of βs and larger positive bias of λ;
- For equations in difference, an increased equation heterogeneity, ${\sigma_\alpha}^2$, or, an increased signal heterogeneity, ${\sigma_\chi}^2$, leads to smaller negative bias of βs and smaller negative bias of λ.

Issue 4: The impact on estimator bias of changed panel size and time-series length

- N is reduced from 100 to 50 and T is reduced from 10 to 8 and further to 6 for both large and small signal spread $\left(\sigma_{\psi}{}^{2} = 1.0, 0.5\right)$:
  - The sensitivity of the means are dramatic within the range;
  - For equation in level, the bias tends to be smaller for N=100 than for N=50, however, this result does not hold for the equation in differences;
  - Neither does the bias tend to vary monotonically with T for a fixed N.

FDI Impact on GDP Revisited

Motivation of using our approach

- Razin and Sadka (2012) noted that since different countries have different recording and accounting practices relating to FDI, measurement errors are likely to arise;
- Neuhaus (2006) pointed out that measurement error is a prevalent problem in a transition country when the FDI impact on its economic growth is investigated.

Advantage of using our approach

- Compared to the common used B&B method, our model provides a theoretical justification and criteria to select the appropriate IVs to incorporate the effect of measurement errors;
- As what we will show below, our approach change the conclusions dramatically.

Revisions for the previous findings in the literature

- Previous: Manufacturing FDI has no significant effects on either the Manufacturing GDP or the Service GDP
- Ours: Manufacturing FDI has significant positive contributions on both the Manufacturing GDP and the Service GDP;
- Previous: Service FDI has positive contribution to service GDP but negative effect on manufacturing GDP (spill-over)
- Ours: Service FDI has positive significant contribution to manufacturing GDP.

Data

- A balanced data set of 1572 observations, including 131 countries and 12 years from 1996 to 2010;
- Variables include real GDP per capita in 2005 PPP, FDI and GDP in current USD, the degree of openness, and working age population from WDI, and the political stability index from WGI;

Empirical Specification

$$Y_{it} = \alpha_i + Y_{i,t-1}\lambda + X_{it}\beta + Z_{it}\gamma + \varepsilon_{it},$$

Table 1. Impact of Aggregated FDI on GDP under OLS and System GMM (Blundell and Bond (1998))

	Fixed Effects OLS		Dynamic GMM	
	Est.	S.e.	Est.	S.e.
ln(GDP(-1))	0.846	0.010	0.650	0.061
FDI share	$\boxed{0.029}$	0.013	$\boxed{0.017}$	0.013
Political Stability	0.022	0.004	0.031	0.011
Openness	0.000	0.000	0.001	0.000
Working Age Population Share	0.006	0.001	0.029	0.008
Sargan Test:			123.52	(p=0.1784)
Second-order Autocorrelation Test:			-0.4572	(p=0.3237)
Wald Test for Coefficients:			1062.84	(p=0.0000)

Motivated by the results from the Sargan test and tests for second-order disturbance autocorrelation test the IVs used are all values of X and Z lagged 5 periods.

Table 2. Impact of Aggregated FDI on GDP

GMM based on q-IVs. $N = 131$, $T = 12$ (years 1996,1998,2000,2002–2010)

	$(N_\xi, N_\eta, N_\nu) =$					
	(4,0,0)	(4,1,0)	(4,2,0)	(3,0,0)	(3,1,0)	(2,0,0)
	Equation in levels					
Est.	0.0388	0.0361	0.0288	0.0408	0.0392	0.0302
S.e.	0.0276	0.0274	0.0213	0.0280	0.0295	0.0209
PS-R^2	0.6674	0.6144	0.5572	0.6346	0.5779	0.5903
$\mathcal{J}$-test	1.0000	0.9690	0.0921	0.9993	0.2033	0.4836
	Equation in differences					
Est.	0.0198	0.0271	0.0235	0.0198	0.0250	0.0236
S.e.	0.0136	0.0221	0.0205	0.0149	0.0220	0.0176
PS-R^2	0.4216	0.4322	0.4097	0.4227	0.4301	0.4179
$\mathcal{J}$-test	1.0000	0.9788	0.0606	1.0000	0.2424	0.4836

Table 3. Impact of Aggregated FDI on GDP for Asian Developing Countries

GMM estimates based on q-IVs only. $N = 24$, $T = 9$ (years 2002–2010)

	$(N_\xi, N_\eta, N_\nu) =$					
	(2,0,0)	(2,1,0)	(3,2,1)	(3,1,0)	(4,1,0)	(4,2,1)
	Equation in levels					
Est.	0.1881	0.1920	0.1600	0.1650	0.1674	0.1312
S.e.	0.0884	0.0639	0.1600	0.0644	0.0595	0.0806
PS-R^2	0.8390	0.7403	0.7214	0.8142	0.8580	0.7808
$\mathcal{J}$-test	1.0000	0.9197	0.6303	1.0000	1.0000	0.9995
	Equation in differences					
Est.	-0.3385	-0.5877	-0.5224	-0.4716	-0.3296	-0.2629
S.e.	0.2971	0.4061	0.4555	0.3618	0.3141	0.3612
PS-R^2	0.5708	0.6053	0.5949	0.5930	0.5803	0.5777
$\mathcal{J}$-test	1.0000	0.9197	0.6303	1.0000	1.0000	0.9994

- Jointly judged by PS-R^2 *and J* test, we found the constellation of (4,0,0) yields the best estimate for the worldwide case whereas (4,1,0) gives the best for Asian Developing Countries;
- For the worldwide case, the effect of FDI is much higher than results presented in Table 1;
- For Asian Developing Countries, the contribution of FDI on GDP is much higher than the worldwide case (about four times).

Table 4. Impact of Manufacturing FDI

GMM ESTIMATES BASED ON q-IVS. $N=32$, $T=8$ (YEARS 2002–2009)

	$(N_\xi, N_\eta, N_\nu) =$ (2,0,0)	(2,1,0)	(3,2,1)	(3,1,0)	(4,1,0)	(4,2,1)
A. On Manufacturing GDP						
Equation in levels						
Est.	0.2672	0.3961	0.7127	0.6713	0.6003	1.3858
S.e.	0.2001	0.3051	0.5368	0.3292	0.3469	0.7005
PS-R^2	0.6444	0.4260	0.5265	0.6063	0.6874	0.6401
$\mathcal{J}$-test	0.9999	0.2321	0.0790	0.9828	0.9999	0.6137
Equation in differences						
Est.	0.2946	0.2046	0.0152	0.1646	0.2244	0.3051
S.e.	0.1703	0.2089	0.4010	0.1753	0.1558	0.2154
PS-R^2	0.6910	0.6851	0.6773	0.6742	0.6822	0.6771
$\mathcal{J}$-test	0.9999	0.2321	0.0799	0.9837	0.9999	0.6137
B. On Service GDP						
Equation in levels						
Est.	0.1684	0.5018	0.6681	0.5159	0.3614	0.9706
S.e.	0.1249	0.3660	0.3061	0.2227	0.2100	0.4970
PS-R^2	0.6422	0.4232	0.5231	0.6039	0.6846	0.6359
$\mathcal{J}$-test	0.9999	0.2321	0.0475	0.9828	0.9999	0.6137
Equation in differences						
Est.	-0.1702	-0.1452	-0.1613	-0.1580	-0.1759	-0.0125
S.e.	0.3907	0.3783	0.4483	0.3475	0.3550	0.3659
PS-R^2	0.7372	0.7281	0.7346	0.7252	0.7261	0.7325
$\mathcal{J}$-test	0.9995	0.2321	0.0401	0.9830	0.9999	0.6137

- Jointly judged by PS-R^2 *and J* test, the constellation of (4,1,0) yields the best estimates ;
- Contribution of manufacturing FDI on Manufacturing GDP is almost two times than its impact on Service GDP;

Table 5. Impact of Service FDI

GMM ESTIMATES BASED ON q-IVs. $N=29$, $T=8$ (YEARS 2002–2009)

	$(N_\xi, N_\eta, N_\nu) =$ (2,0,0)	(2,1,0)	(3,2,1)	(3,1,0)	(4,1,0)	(4,2,1)
	A. ON SERVICE GDP					
	Equation in levels					
Est.	0.0780	0.1851	0.1993	0.1127	0.0356	0.0280
S.e.	0.0404	0.0872	0.1342	0.0839	0.0745	0.0829
PS-R^2	0.6460	0.4319	0.5269	0.6214	0.7016	0.6530
$\mathcal{J}$-test	0.9999	0.3609	0.0906	0.9944	0.9999	0.7523
	Equation in differences					
Est.	0.2686	0.2766	0.1781	0.2607	0.2379	0.2407
S.e.	0.0926	0.0860	0.0748	0.0827	0.0810	0.0653
PS-R^2	0.7605	0.7418	0.7377	0.7337	0.7302	0.7440
$\mathcal{J}$-test	1.0000	0.3609	0.0882	0.9944	0.9999	0.7523
	B. ON MANUFACTURING GDP					
	Equation in levels					
Est.	0.0481	0.1985	0.4475	0.1594	0.0842	0.2140
S.e.	0.0614	0.1045	0.1846	0.0774	0.0612	0.0990
PS-R^2	0.6470	0.4326	0.5287	0.6221	0.7034	0.6558
$\mathcal{J}$-test	0.9999	0.3609	0.0934	0.9944	0.9999	0.7523
	Equation in differences					
Est.	0.1489	0.2186	0.2510	0.1689	0.1364	0.1989
S.e.	0.0674	0.0921	0.0812	0.0659	0.0655	0.0822
PS-R^2	0.6951	0.6840	0.6898	0.6975	0.6982	0.6952
$\mathcal{J}$-test	1.0000	0.3609	0.1006	0.9938	1.0000	0.7523

- Jointly judged by PS-R^2 *and J* test, the constellation of (2,0,0) yields the best estimates , which implies the length of memory in the latent service GDP is shorter than that in the latent manufacturing GDP;
- Contribution of service FDI on service GDP is almost two times than its impact on Manufacturing GDP. However, the magnitude is much smaller than the impact of the manufacturing GDP;
- Equation in differences is better than equation in levels.

The End