Large-*N* and Large-*T* Properties of Dynamic Panel GMM Estimators When Data Are Not Mean Stationary

Seung C. Ahn
Arizona State University

Na Wang
Hofstra University

AEA-KAEA Session
January 6, 2018

1. Motivation

- A simple dynamic panel model

  - $y_{it} = \delta y_{i,t-1} + (1-\delta)\alpha_i + \varepsilon_{it}$, where the ε_{it} are *i.i.d.*

    - Equivalent to $y_{it} = \alpha_i + u_{it}; u_{it} = \delta u_{i,t-1} + \varepsilon_{it}$.

  - Initial Obs.: $y_{i0} = \gamma\alpha_i + u_{i0}$.

    - $\mathrm{E}(y_{ii} \mid \alpha_i) = [1-(1-\gamma)\delta^{t-1}]\alpha_i$

    - If $\gamma = 1$, $\mathrm{E}(y_{ii} \mid \alpha_i) = \alpha_i \;\rightarrow\; y_{it}$ is said to be mean stationary.

- Popular GMM estimation
  - Legitimate moment conditions:
    - $\mathrm{E}[y_{i,t-2}(\Delta y_i - \delta \Delta y_{i,-1})] = \mathrm{E}(y_{i,t-2}\Delta\varepsilon_{it}) = 0$, and many others.
    - GMM using these moment conditions is called GMM using level instruments to differenced equations.
      - Known to be unreliable if δ is close to one.

- Popular GMM estimation

    - Legitimate moment conditions if $\gamma = 1$:

        - $\mathrm{E}[\Delta y_{i,t-1}(y_i - \delta y_{i,-1})] = \mathrm{E}(\Delta y_{i,t-1}[(1-\delta)\alpha_i + \varepsilon_{it}]) = 0$,
          because $\Delta y_{i,t-1} = \Delta u_{i,t-1}$ is uncorrelated with α_i.

        - GMM using these moment conditions is called GMM using differenced instruments to level equations.

        - We refer as the "MS" GMM estimators to the GMM estimators using these moment conditions.

- Known facts for the data with large N and small T
  - If $\gamma \neq 1$, MS GMM estimators are inconsistent.
  - If $\gamma = 1$, there is efficiency gain by using MS moment conditions.
  - MS GMM estimators are particularly useful when δ is near one.

- What if both N and T are large?
  - Are MS estimators still consistent even if $\gamma \neq 1$?

    Why possible? $y_{it} = [1-(\gamma-1)\delta^{t-1}]\alpha_i + u_{it} \to \alpha_i + u_{it}$ as $t \to \infty$.
  - Is there efficiency gain?
  - What if δ is close to one or one?
- These questions motivate this paper.
  - Try to provide some insights.

- What this paper does.

  - Consider three estimators

    - Anderson-Hsiao Estimator

      - GMM using $g_{AH,iT} = \Sigma_{t=1}^{T} y_{i,t-2}(\Delta y_{it} - \delta \Delta y_{i,t-1})$.
      - Consistent even if $\gamma \neq 1$.
      - Hsiao and Zhang (2015) and Phillips (2015)

    - MS1 Estimator = GMM using $g_{MS1,iT} = \Sigma_{t=2}^{T} \Delta y_{i,t-1}(y_{it} - d y_{i,t-1})$.
    - MS2 Estimator = GMM using both $g_{AH,iT}(d)$ and $g_{MS1,iT}(d)$.

- What this paper does.
  - Consider three cases
    - $|\delta| < 1$.
    - $\delta = 1 - \dfrac{\psi}{T^c}, \;\; \psi, c > 0$.
    - $\delta = 1$ (turns out to be equivalent to the case of $c > 3/2$).

- Why MS1? How representative it is for MS estimation?

  - All available linear moment functions even if $\gamma \neq 1$:

    $$\mathrm{E}(y_{is}(\Delta y_{it} - \delta \Delta y_{i,t-1})) = 0, \quad s < t-1.$$

  - All available non-redundant linear moment functions when $\gamma = 1$:

    $$\mathrm{E}(\Delta y_{i,t-1}(y_i - \delta y_{i,t-1})) = 0 \;\text{ for }\; t = 2, 3, ..., T.$$

  - MS1 uses $\mathrm{E}(\Sigma_{t=2}^{T} \Delta y_{i,t-1}(y_i - \delta y_{i,t-1})) = 0$.

2. Preliminaries

Basic Assumption (BA):

(i) The random vectors $(y_{i0}, \alpha_i)'$ are *i.i.d.* over different *i* and have zero means and finite second moments.

$$\mathrm{E}(\alpha_i^2) = \sigma_\alpha^2 < \infty; \;\; y_{i0} = \gamma\alpha_i + u_{i0}; \;\; \mathrm{E}(u_{i0} \mid \alpha_i) = 0, \;\; \mathrm{E}(u_{i0}^2 \mid \alpha_i) = \sigma_0^2$$

(ii) For all *t*, the ε_{it} are i.i.d. over different *i* and *t*, and independent from the $(y_{i0}, \alpha_i)'$, and have finite moments up to 4th order, with $\mathrm{E}(\varepsilon_{it}) = 0$ and $\mathrm{E}(\varepsilon_{it}^2) = \sigma_\varepsilon^2 < \infty$.

(iii) As $N, T \to \infty$, $N/T \to \rho < \infty$ (Can be relaxed.)

- Link between AH and MS1 moment functions

- $$g_{MS1,iT}(d) \equiv \Sigma_{t=2}^{T} \Delta y_{i,t-1}(y_{it} - dy_{i,t-1})$$
$$= [(y_{i,T-1}y_{iT} - y_{i0}y_{i1}) - (y_{i,T-1}^2 - y_{i0}^2)d]$$
$$-\Sigma_{t=2}^{T} y_{i,t-2}(\Delta y_{it} - d\Delta y_{i,t-1})$$
$$\equiv m_{iT}(d) - g_{AS,iT}(d)$$

- $$\frac{1}{N^{1/2}T^{1/2}} \Sigma_{I=1}^{N} g_{MS1,iT}(\delta)$$
$$= \frac{1}{N^{1/2}T^{1/2}} \Sigma_{i=1}^{N} m_{iT}(\delta) - \frac{1}{N^{1/2}T^{1/2}} \Sigma_{i=1}^{N} g_{AS,iT}(\delta)$$

- Estimators
  - AH estimator: $$\hat{\delta}_{AH} = \frac{\Sigma_{i=1}^{N}\Sigma_{t=2}^{T} y_{i,t-2}\Delta y_{it}}{\Sigma_{i=1}^{N}\Sigma_{t=2}^{T} y_{i,t-2}\Delta y_{i,t-1}}$$
  - MS1 estimator: $$\hat{\delta}_{MS1} = \frac{\Sigma_{i=1}^{N}\Sigma_{t=2}^{T} \Delta y_{i,t-1} y_{it}}{\Sigma_{i=1}^{N}\Sigma_{t=2}^{T} \Delta y_{i,t-1} y_{i,t-1}}$$
  - MS2 estimator: $$\hat{\delta}_{MS2} = \frac{\begin{pmatrix} \Sigma_{i=1}^{N}\Sigma_{t=2}^{T} y_{i,t-2}\Delta y_{i,t-1} \\ \Sigma_{i=1}^{N}\Sigma_{t=2}^{T} \Delta y_{i,t-1} y_{i,t-1} \end{pmatrix}' \hat{V}_{MS2}^{-1} \begin{pmatrix} \Sigma_{i=1}^{N}\Sigma_{t=2}^{T} y_{i,t-2}\Delta y_{it} \\ \Sigma_{i=1}^{N}\Sigma_{t=2}^{T} \Delta y_{i,t-1} y_{it} \end{pmatrix}}{\begin{pmatrix} \Sigma_{i=1}^{N}\Sigma_{t=2}^{T} y_{i,t-2}\Delta y_{i,t-1} \\ \Sigma_{i=1}^{N}\Sigma_{t=2}^{T} \Delta y_{i,t-1} y_{i,t-1} \end{pmatrix}' \hat{V}_{MS2}^{-1} \begin{pmatrix} \Sigma_{i=1}^{N}\Sigma_{t=2}^{T} y_{i,t-2}\Delta y_{i,t-1} \\ \Sigma_{i=1}^{N}\Sigma_{t=2}^{T} \Delta y_{i,t-1} y_{i,t-1} \end{pmatrix}}$$

$$\hat{V}_{MS2} = \frac{1}{N(T-1)}\Sigma_{i=1}^{N} \begin{pmatrix} (g_{AH,iT}(\tilde{\delta}))^2 & g_{AH,iT}(\tilde{\delta}) g_{MS1,iT}(\tilde{\delta}) \\ g_{AH,iT}(\tilde{\delta}) g_{MS1,iT}(\tilde{\delta}) & (g_{MS1,iT}(\tilde{\delta}))^2 \end{pmatrix}$$

$$J_{MS} = \frac{1}{N(T-1)} \begin{pmatrix} \Sigma_{i=1}^{N} g_{AH,iT}(\hat{\delta}_{MS2}) \\ \Sigma_{i=1}^{N} g_{MS1}(\hat{\delta}_{MS2}) \end{pmatrix}' \hat{V}_{MS2}^{-1} \begin{pmatrix} \Sigma_{i=1}^{N} g_{AH,iT}(\hat{\delta}_{MS2}) \\ \Sigma_{i=1}^{N} g_{MS1}(\hat{\delta}_{MS2}) \end{pmatrix}.$$

3. Asymptotic Results When $|\delta| < 1$

Theorem 1: Under BA with $|\delta| < 1$, as $N, T \to \infty$,

$$(NT)^{1/2}\left(\hat{\delta}_{AH} - \delta\right) = X + o_p(1);$$

$$(NT)^{1/2}(\hat{\delta}_{MS1} - \delta) = \rho^{1/2}\beta_{MS1}(\delta,\gamma) + X + o_p(1);$$

$$(NT)^{1/2}(\hat{\delta}_{MS2} - \delta) = \rho^{1/2}\beta_{MS2}(\delta,\gamma) + X + o_p(1);$$

$$J_{MS} = \frac{\rho}{2(1+\delta)}\left(\beta_{MS2} - \beta_{MS1}\right)^2 + \frac{\left(N^{1/2}(1-\delta)(1-\gamma)\sigma_\alpha^2 + H / \sigma_H\right)^2}{(1-\delta)^2(1-\gamma)^2\sigma_\alpha^2 / \sigma_H^2 + 1} + o_p(1),$$

where $X \sim N(0, 2(1+\delta))$

4. Asymptotic Results When $\delta = 1 - \psi T^{-c}$

Theorem 2: Under BA and LAU, the following holds as $N, T \to \infty$.

$$T^{1/2}(\hat{\delta}_{AH} - \delta) \approx \begin{cases} 2C, & \text{if } c > \dfrac{3}{2}; \\ N\left(0, \dfrac{2}{\rho \xi_{AH,1}^{2}}\right) + 2 \times C, & \text{if } c = \dfrac{3}{2}; \\ N\left(0, \dfrac{2}{\rho \psi^{2}\left({}^{2}\mathrm{E}\left[\int_0^1 (W(r))^2 dr\right]\right)^{2}}\right), & \text{if } 1 < c < \dfrac{3}{2}; \end{cases}$$

$$(NT)^{1/2}\left(\hat{\delta}_{AH}-\delta\right)\approx\begin{cases} N\left(0,\dfrac{\left(1+\mathrm{E}\left[\left(J(1)\right)^2\right]\right)}{\psi^2\left(\mathrm{E}\left[\int_0^1 (J(r))^2 dr\right]\right)^2}\right), & \text{if } c=1; \\ N(0,4), & \text{if } c<1, \end{cases}$$

$$(NT)^{1/2}\left(\hat{\delta}_{MS1}-\delta\right)\approx\begin{cases} N(0,1), & \text{if } c>1; \\ N\left(0,\dfrac{1}{a_1}\right), & \text{if } c=1; \\ N(0,4)+o_p(1), & \text{if } c<1, \end{cases}$$

where $a_1=\left(\mathrm{E}((J(1))^2)-\psi\mathrm{E}\left[\int_0^1 (J(r))^2 dr\right]\right)^2.$

$$
(NT)^{1/2}\left(\hat{\delta}_{MS2}-\delta\right)\approx\begin{cases} N\left(0,\dfrac{1}{2}\right), & \text{if } c>1; \\ N\left(0,\dfrac{1}{a_1+a_2}\right), & \text{if } c=1; \\ N(0,4), & \text{if } c<1, \end{cases}
$$

where $a_2 = \mathrm{E}((J(1))^2)$.

$J_{MS} \approx \chi^2(1) + o_p(1)$ (convergence rate is $(NT)^{1/2}$) if $c \geq 1$

$$
J_{MS} \approx \left(\frac{\rho^{1/2}}{(T-1)^{(3c-1)/2}} \frac{\psi^{3/2}}{2^{1/2}} (1-\gamma) \frac{\sigma_\alpha^2}{\sigma_\varepsilon^2} + H_3 \right)^2 \text{ if } c<1,
$$

where $H_3 \sim N(0,1)$.

Remarks

(1) AH estimator is normal if $c < 3/2$.

(2) MS1 and MS2 are always asymptotically unbiased and normal.

(3) When $1 \leq c < 3/2$, MS2 is more efficient than AH and MS1.
When $c < 1$, AH, MS1 and MS2 are asymptotically equivalent

(4) $J_{MS} \approx \chi^2(1)$ if $c > 1/3$; $\to \infty$, if $c < 1/3$.

(5) *t*-tests for the unit root hypothesis (from AH, MS1 and MS2) are consistent if $c < 1$.

(6) MS1 and MS2 can be asymptotically biased if $\rho = \infty$.

5. Asymptotic Results When $\delta = 1$

$$T^{1/2}(\hat{\delta}_{AH} - \delta) = 2C + o_p(1);$$

$$(NT)^{1/2}\left(\hat{\delta}_{MS1} - \delta\right) = N(0,1) + o_p(1);$$

$$(NT)^{1/2}\left(\hat{\delta}_{MS2} - \delta\right) = N\left(0, \frac{1}{2}\right) + o_p(1);$$

$$J_{MS} = \chi^2(1) + o_p(1).$$

Remark

J_{MS} is $\chi^2(1)$ when data follow unit root processes, as well as when data are mean stationary.

6. Simulation Results

$$\sigma_{\alpha}^{2} = 4; \ \ \sigma_{0}^{2} = \mathrm{var}(y_{i0}) = 4; \ \ \sigma_{\varepsilon}^{2} = 1.$$

Table 1: Root Mean Square Errors ($\gamma = 1$)

		Estimator	δ=0.5	δ=0.8	δ=0.9	δ=0.95	δ =0.99	δ =1.0
$N = 100$	$T = 10$	AH	0.066	0.088	**0.125**	**0.454**	**67.69**	**80.63**
		MS2	**0.064**	**0.067**	**0.058**	**0.045**	**0.033**	**0.030**
	$T = 50$	AH	0.027	0.031	0.034	0.040	**0.185**	**41.02**
		MS2	0.026	0.028	0.028	0.025	0.015	0.011
	$T = 100$	AH	0.018	0.020	0.021	0.023	0.041	**6.197**
		MS2	0.018	0.019	0.019	0.018	0.011	0.007
	$T = 200$	AH	0.012	0.014	0.014	0.015	0.020	**6.424**
		MS2	0.012	0.013	0.014	0.013	0.010	0.005
	$T = 400$	AH	0.009	0.009	0.010	0.010	0.012	**1.438**
		MS2	0.009	0.009	0.010	0.009	0.008	0.004

Table 2: Sizes of *t*-Tests with MS2 Estimator ($\gamma = 1$)

		Estimator	$\delta = 0.5$	$\delta = 0.8$	$\delta = 0.9$	$\delta = 0.95$	$\delta = 0.99$	$\delta = 1.0$
$N = 100$	$T = 10$	AH	0.062	0.054	0.040	0.018	0.003	0.002
		MS2	0.068	0.058	0.051	0.050	0.055	**0.055**
	$T = 50$	AH	0.066	0.068	0.060	0.055	0.030	0.000
		MS2	0.062	0.053	0.056	0.064	0.072	**0.073**
	$T = 100$	AH	0.049	0.058	0.055	0.059	0.043	0.000
		MS2	0.056	0.056	0.056	0.056	0.038	**0.040**
	$T = 200$	AH	0.058	0.049	0.045	0.047	0.041	0.000
		MS2	0.060	0.056	0.047	0.051	0.051	**0.059**
	$T = 400$	AH	0.044	0.050	0.047	0.045	0.037	0.001
		MS2	0.046	0.050	0.048	0.048	0.056	**0.063**

MS2, reasonably well sized

Table 3: Power of the *t*-Tests for the Unit Root Hypothesis ($\gamma = 1$)

		Estimator	$\delta = 0.8$	$\delta = 0.9$	$\delta = 0.95$	$\delta = 0.99$
$N = 100$	$T = 10$	AH	0.632	0.139	0.049	0.003
		MS2	0.844	0.432	0.211	0.061
	$T = 50$	AH	1.000	0.860	0.266	0.030
		MS2	1.000	0.967	0.565	0.118
	$T = 100$	AH	**1.000**	**0.997**	**0.564**	0.054
		MS2	**1.000**	**1.000**	**0.791**	**0.140**
	$T = 200$	AH	1.000	1.000	0.917	0.082
		MS2	1.000	1.000	0.967	0.177
	$T = 400$	AH	**1.00**	**1.000**	**0.999**	0.137
		MS2	**1.00**	**1.000**	**0.999**	**0.245**

Table 5: Root Mean Square Errors

Table 4: Size of the *J*-Test for the Mean Stationarity Hypothesis ($\gamma = 1$)

		$\delta = 0.5$	$\delta = 0.8$	$\delta = 0.9$	$\delta = 0.95$	$\delta = 0.99$	$\delta = 1.0$
$N = 100$	$T = 5$	0.053	0.040	0.040	0.041	0.047	**0.048**
	$T = 10$	0.061	0.053	0.051	0.049	0.043	**0.047**
	$T = 50$	0.050	0.049	0.053	0.058	0.052	**0.055**
	$T = 100$	0.050	0.055	0.054	0.047	0.046	**0.045**
	$T = 200$	0.051	0.052	0.047	0.041	0.044	**0.043**
	$T = 400$	0.036	0.049	0.051	0.062	0.044	**0.037**

Reasonably well sized.

Table 5: Root Mean Square Errors ($\gamma = 4$)

		Estimator	$\delta = 0.5$	$\delta = 0.8$	$\delta = 0.9$	$\delta = 0.95$	$\delta = 0.99$	$\delta = 1.0$
$N = 100$	$T = 10$	AH	**0.023**	0.027	0.038	0.063	9.864	**10.89**
		MS2	**0.147**	0.066	0.036	0.036	0.083	**0.039**
	$T = 50$	AH	0.018	0.020	0.021	0.023	0.055	9.627
		MS2	0.082	0.042	0.026	0.021	0.031	0.014
	$T = 100$	AH	0.014	0.016	0.016	0.017	0.028	2.276
		MS2	0.049	0.028	0.019	0.017	0.021	0.009
	$T = 200$	AH	0.011	0.012	0.012	0.013	0.017	28.51
		MS2	0.027	0.017	0.013	0.013	0.014	0.006
	$T = 400$	AH	**0.008**	0.009	0.009	0.010	0.011	**2.688**
		MS2	**0.014**	0.011	0.010	0.010	0.010	**0.004**

Table 6: Size of the *t*-Test with the MS2 Estimator ($\gamma = 4$)

		$\delta = 0.5$	$\delta = 0.8$	$\delta = 0.9$	$\delta = 0.95$	$\delta = 0.99$	$\delta = 1.0$
$N = 100$	$T = 10$	**1.000**	**0.999**	**0.646**	**0.073**	**0.029**	**0.035**
	$T = 50$	0.997	0.774	0.322	0.098	0.057	0.058
	$T = 200$	0.613	0.235	0.094	0.052	0.043	0.055
	$T = 400$	**0.334**	**0.128**	**0.077**	**0.064**	**0.044**	**0.058**

Table 7: Power of *J* Test for the Mean Stationarity Hypothesis ($\gamma = 4$)

		$\delta = 0.5$	$\delta = 0.8$	$\delta = 0.9$	$\delta = 0.95$	$\delta = 0.99$	
$N = 100$	$T = 10$	**1.000**	**0.709**	**0.130**	**0.052**	**0.038**	
	$T = 50$	1.000	0.837	0.325	0.126	0.046	
	$T = 200$	1.000	0.798	0.287	0.096	0.037	
	$T =400$	**1.000**	**0.780**	**0.278**	**0.084**	**0.051**	

Main findings from simulations

- AH performs well even if δ is close to one.
  - When δ is not close to one, little gain by MS2.
  - When data are not MS and δ is not close to one, the *t*-test from MS2 over-rejects correct hypothesis, especially when $N \gg T$.
  - The MS assumption should be routinely tested.

- MS2 is useful when δ is close to one.
  - The unit root hypothesis can be tested by usual *t*-test.
  - Bias is small when δ is close to one or T is much larger than *N*.

8. Concluding Remarks

- While limited, our findings can have three general implications.

- GMM estimators using the MS conditions are asymptotically biased if data are not mean-stationarity and data are only mildly persistent. The mean stationarity assumption should be routinely tested using the over-identifying restriction statistics.

- Contrary to common belief, GMM estimators not implementing MS conditions can produce reliable inferences even if data are of near unit roots.

- The MS moment conditions would be most useful for the analysis of persistent data.