Economics Bulletin

Volume 32, Issue 1

The tests for the level moment conditions: GMM estimation in a linear dynamic panel data model

Takuya Hasebe
Graduate Center, City University of New York

Abstract

This paper compares the power of alternative tests for the level moment conditions in GMM estimation of a linear dynamic panel data model. The test statistic calculated conventionally can take on a negative value in finite samples even though it cannot be asymptotically. A straightforward modification makes the test statistic nonnegative. Monte Carlo simulations show that the test can gain power from the modification.

I would like to acknowledge Wim Vijverberg and Catherine Lau for their valuable comments. All errors are my own.
Citation: Takuya Hasebe, (2012) "The tests for the level moment conditions: GMM estimation in a linear dynamic panel data model", *Economics Bulletin*, Vol. 32 No. 1 pp. 412-420.
Contact: Takuya Hasebe - thasebe@gc.cuny.edu.
Submitted: October 04, 2011. **Published:** January 27, 2012.

1 Introduction

In the estimation of dynamic panel data models, the first-difference GMM (DIF-GMM) estimator by Arellano and Bond (1991) and the system GMM (SYS-GMM) estimator by Arellano and Bover (1995) and Blundell and Bond (1998) have been used widely. The latter estimator exploits moment conditions additional to those of the former estimator. The validity of these additional moment conditions, which are referred to as the level moment conditions, requires mean-stationarity of initial observations. Although the level moment conditions lead the SYS-GMM estimator to better finite-sample properties than the DIF-GMM estimator, the violation of the level conditions results in inconsistency of the SYS-GMM estimator. As Roodman (2009) cautions, there is a risk of underappreciating the initial condition. The standard way to test these additional moment conditions is to take a difference between the Sargan test statistics of the SYS-GMM and DIF-GMM estimators.

However, little is known about the finite sample properties of the test, especially the power properties.[1] Furthermore, the test statistic can be negative in finite samples even though it has an asymptotic χ^2 distribution. Indeed, a nontrivial portion of Monte Carlo simulations by Blundell and Bond (2000) resulted in negative values.

This paper considers a modification of the test statistic to avoid a negative value. Monte Carlo simulations illustrate that the modified test performs better than the conventional test in finite samples.

2 Model

Consider an AR(1) panel data model:

$$y_{i,t} = \alpha y_{i,t-1} + u_{i,t}, \qquad i = 1, \dots, N; t = 2, \dots, T,$$

where $|\alpha| < 1$ and $u_{i,t}$ is an unobservable error: $u_{i,t} = \eta_i + \varepsilon_{i,t}$. Assume that $\eta_i \sim iid(0, \sigma_\eta^2)$ and $\varepsilon_{i,t} \sim iid(0, \sigma_\varepsilon^2)$. The initial observation is generated as:

$$y_{i,1} = \frac{\delta}{1-\alpha}\eta_i + \omega_{i,1}, \qquad i = 1, \dots, N,$$

where δ is a constant and $\omega_{i,1} \sim iid(0, \sigma_\omega^2)$. The DIF-GMM estimator exploits the $(T-1)(T-2)/2$ moment conditions:

$$E\left(Z_{di}{}'\Delta u_i\right) = 0, \tag{1}$$

where

$$Z_{di} = \begin{pmatrix} y_{i,1} & 0 & 0 & \dots & 0 & \dots & 0 \\ 0 & y_{i,1} & y_{i,2} & \dots & 0 & \dots & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot & \dots & \cdot \\ 0 & 0 & 0 & \dots & y_{i,1} & \dots & y_{i,T-2} \end{pmatrix} \text{ and } \Delta u_i = \begin{pmatrix} \Delta u_{i,3} \\ \vdots \\ \Delta u_{i,T} \end{pmatrix}.$$

[1] In dynamic panel data models, Bowsher (2002) examines the power of the Sargan test in the presence of serial correlation, and Dahlberg et al. (2008) examine it with measurement errors in data. However, neither of them considers the test specific to the level moment conditions. Only Roodman (2009) pays special attentions to the level moment conditions.

These moment conditions do not require any restriction on δ.

If the initial observation is mean-stationary[2] such that $\delta = 1$, an additional $(T-2)$ moment conditions are available:

$$E\left(Z_{li}{}'u_i\right) = 0,$$

where

$$Z_{li} = \begin{pmatrix} \Delta y_{i,2} & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & \Delta y_{i,T-1} \end{pmatrix} \text{ and } u_i = \begin{pmatrix} u_{i,3} \\ \vdots \\ u_{i,T} \end{pmatrix}.$$

These additional conditions are referred to as the level moment conditions. The SYS-GMM estimator exploits these conditions in addition to (1). Defining

$$Z_{si} = \begin{pmatrix} Z_{di} & 0 \\ 0 & Z_{li} \end{pmatrix} \text{ and } \nu_i = \begin{pmatrix} \Delta u_i \\ u_i \end{pmatrix},$$

the moment conditions in the SYS-GMM estimator are expressed as

$$E\left(Z_{si}{}'\nu_i\right) = E\begin{pmatrix} Z_{di}{}'\Delta u_i \\ Z_{li}{}'u_i \end{pmatrix} = 0 \tag{2}$$

It is well-documented that, in finite samples, the SYS-GMM estimator performs better than the DIF-GMM estimator, which suffers from the weak instruments problem. Therefore, it is becoming more popular in empirical applications. If $\delta \neq 1$, however, the level moment conditions are invalid, and it leads the SYS-GMM estimator to be inconsistent although the DIF-GMM estimator remains consistent. Therefore, it is important to test the validity of the level moment conditions.

3 The Overidentifying Restrictions Tests

As both the DIF-GMM and SYS-GMM estimators are overidentified, the validity of the moment conditions can be tested. The standard test is the Sargan test. The test statistic for the DIF-GMM estimator is given by

$$\mathrm{Sar}_d(\hat{\alpha}_{d1}) = \frac{1}{N}\widehat{\Delta u}' Z_d \hat{S}_d(\hat{\alpha}_{d1})^{-1} Z_d{}'\widehat{\Delta u},$$

where Z_d is the matrix stacking individuals' instrument matrices and $\widehat{\Delta u}$ is the vector stacking individuals' residual vectors from a two-step DIF-GMM estimator. $\hat{S}_d(\hat{\alpha}_{d1})$ is the consistent estimator of the covariance matrix of the sample moments in the DIF-GMM estimator, which is given by

$$\hat{S}_d(\hat{\alpha}_{d1}) = \frac{1}{N}\sum_{i}^{N} Z_{di}'\widehat{\Delta u_i}\widehat{\Delta u_i}' Z_{di}, \tag{3}$$

[2]If $\sigma_\omega^2 = \sigma_\varepsilon^2/(1-\alpha^2)$, the series is covariance-stationary. The SYS-GMM estimator does not require covariance-stationarity.

where $\widehat{\Delta u_i}$ is the residual vector from a one-step DIF-GMM estimator, $\hat{\alpha}_{d1}$. Similarly, the Sargan test statistic for the SYS-GMM estimator is given by

$$\text{Sar}_s(\hat{\alpha}_{s1}) = \frac{1}{N}\widehat{\nu}' Z_s \hat{S}_s(\hat{\alpha}_{s1})^{-1} Z_s{}'\widehat{\nu},$$

where $\hat{S}_{s1}$ is the consistently estimated covariance of the sample moments in the SYS-GMM estimator. It is estimated by

$$\hat{S}_s(\hat{\alpha}_{s1}) = \frac{1}{N}\sum_i^N Z_{si}{}'\widehat{\nu}_i\widehat{\nu}_i' Z_{si}, \tag{4}$$

where $\hat{\nu}_i$ is the residual vector from a one-step SYS-GMM estimator, $\hat{\alpha}_{s1}$.

In GMM estimation, the subset of the moment conditions can be tested by taking a difference of the Sargan test statistics from two GMM estimators: the one using the whole set of the moment conditions and the other excluding the moment conditions to be tested. As it is clear from (2), the DIF-GMM estimator exploits the subset of the moment conditions in the SYS-GMM estimator; the level moment conditions are excluded. The test is called a Difference-Sargan (Dif-Sar) test and is calculated as

$$\text{Dif-Sar} = \text{Sar}_s(\hat{\alpha}_{s1}) - \text{Sar}_d(\hat{\alpha}_{d1}).$$

This test is specific to the validity of the level moment conditions. It has an asymptotic χ^2 distribution with degrees of freedom equal to the number of the level moment conditions, $(T-2)$. This is the test reported conventionally.

However, in finite samples, this test statistic can be negative, especially when the series is persistent (Blundell and Bond, 2000). Although it has not been considered in the literature, it is possible to avoid this problem. To obtain a nonnegative test statistic, use the same one-step estimator to estimate the covariance matrices of the moments of the DIF- and SYS-GMM estimators. Then, take a difference of Sargan test statistics from the same one-step estimator, whether this estimator is $\hat{\alpha}_{s1}$ or $\hat{\alpha}_{d1}$. Thus, let $C(\hat{\alpha})$ denote the difference Sargan test from a particular one-step estimator $\hat{\alpha}$:

$$C(\hat{\alpha}_{s1}) = \text{Sar}_s(\hat{\alpha}_{s1}) - \text{Sar}_d(\hat{\alpha}_{s1})$$
$$C(\hat{\alpha}_{d1}) = \text{Sar}_s(\hat{\alpha}_{d1}) - \text{Sar}_d(\hat{\alpha}_{d1}).$$

The point here is to use the same estimator $\hat{\alpha}$ to estimate the covariance matrices $\hat{S}_d(\hat{\alpha})$ and $\hat{S}_s(\hat{\alpha})$ in equations (3) and (4).

Asymptotically, Dif-Sar, $C(\hat{\alpha}_{d1})$, and $C(\hat{\alpha}_{s1})$ test statistics are equivalent and have the same limiting χ^2 distribution. However, in finite samples, the values of these tests differ, and Dif-Sar test statistic can be negative while $C(\hat{\alpha}_{d1})$ and $C(\hat{\alpha}_{s1})$ cannot be.[3]

Even though it is straightforward to make a test statistic nonnegative, the applied literature unfortunately relies exclusively on the Dif-Sar test. The question is how the test

[3] To see why, notice that $\hat{S}_d(\hat{\alpha})$ is a submatrix of $\hat{S}_s(\hat{\alpha})$. It is well-known that the use of the submatrix guarantees nonnegativity. See, for example, Hayashi (2000).

Table 1: Relative Frequencies of Rejection of the Level Moment Conditions: $\alpha = 0.8$

	$\hat{\alpha}_{s2}$	$\hat{\alpha}_{d1}$	Dif-Sar	$C(\hat{\alpha}_d)$	$C(\hat{\alpha}_s)$	$C(\alpha)$
$\delta = 0$						
$k = 0$	1.0547 (0.0148)	0.7849 (0.0427)	0.994	0.999	0.999	0.999
2	1.0143 (0.0129)	0.7482 (0.0823)	0.729	0.788	0.876	0.759
5	0.9593 (0.0161)	0.6829 (0.1278)	0.268	0.657	0.464	0.530
10	0.8847 (0.0390)	0.6587 (0.1300)	0.147	0.406	0.252	0.265
20	0.8263 (0.0515)	0.7015 (0.1075)	0.080	0.166	0.132	0.085
40	0.8206 (0.0512)	0.7100 (0.1044)	0.088	0.144	0.133	0.095
$\delta = 2$						
$k = 0$	0.9248 (0.0065)	0.7961 (0.0207)	0.990	1.000	0.997	1.000
2	0.9501 (0.0100)	0.7919 (0.0301)	0.959	0.993	0.976	0.996
5	0.9770 (0.0276)	0.7802 (0.0476)	0.935	0.955	0.966	0.964
10	0.8409 (0.0702)	0.7510 (0.0780)	0.468	0.328	0.518	0.387
20	0.8168 (0.0526)	0.7118 (0.1016)	0.099	0.145	0.150	0.082
40	0.8205 (0.0512)	0.7101 (0.1043)	0.088	0.143	0.134	0.095

Note: The first and second column show the means of two-step SYS-GMM estimates and of one-step DIF-GMM estimates, respectively, with the standard deviations in round brackets. The other columns show relative rejection frequencies of each test. When a test statistic is negative, the decision is non-rejection. The decision of rejection is at the nominal size of 5%. For the parameter values, see the text.

properties vary across the different test statistics. To investigate this issue, I conduct Monte Carlo simulations.

4 Simulation

In this Monte Carlo study, I consider two values of δ that makes the series mean-nonstationary: $\delta = 0$ or 2. I also discard the first k periods of the series. Unless $k = 0$, the initial observation is different from the actual initial starting point of the series. If k is sufficiently large, the series converges to its stationary level so that the initial condition becomes valid. The values of k are 0, 2, 5, 10, 20, and 40. I fix the number of the observed periods at $T = 8$. The problems of the GMM estimators in a dynamic panel data model often arise when the series is persistent. Thus, I set $\alpha = 0.8$. For comparison, I also simulate with $\alpha = 0.4$. The error terms are generated as $\eta_i \sim iidN(0, 1)$, $\varepsilon_{i,t} \sim iidN(0, 1)$, and $\omega_{i,1} \sim iidN\left(0, \sigma_\varepsilon^2/(1 - \alpha^2)\right)$. The number of individuals is $N = 200$. For each combination of the parameters, replication is done 2,000 times.[4]

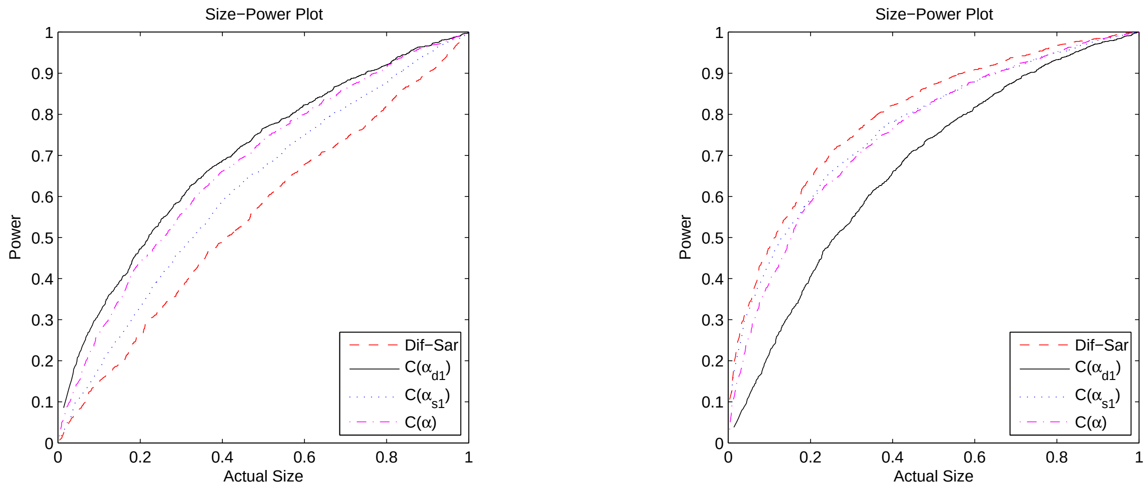
Table 1 summarizes the results.[5] As the first column shows, the invalid initial condition $\delta \neq 1$ causes biases in the SYS-GMM estimator. The biases become gradually smaller as k increases in the case where $\delta = 0$. If $\delta = 2$, the biases intially increase with k but diminish when $k = 10$. When $k = 20$ and $k = 40$, the biases essentially disappear. It is long enough for the series to converge to its stationary level, and initial observations satisfy

[4]All calculations are done with MATLAB using the same random number generating sequences for each setting. I also simulate with $T = 8$, $N = 400$ and different values for σ_ε^2: $\sigma_\varepsilon^2 = 1/4, 1/2, 1, 2, 4$. σ_ω^2 is always set to $\sigma_\omega^2 = \sigma_\varepsilon^2/(1 - \alpha^2)$.

[5] Other results are available upon request.

Figure 1: The Size-Power Plots: $\alpha = 0.8$

(a) $\delta = 0$ and $k = 10$ (b) $\delta = 2$ and $k = 10$



mean-stationarity. However, when k is small, the effect of the invalid initial condition still remains.

The last four columns show the rejection frequencies of each test. For comparison, I also calculate an infeasible test statistic $C(\alpha)$, which uses a true parameter α to estimate the covariance of the moments. The decision to reject is made at the nominal size of 5%. The Dif-Sar test statistic takes on negative values at times while the C test statistic does not by its design. Although both tests can usually reject the invalid moment conditions at $k = 0$, their power declines rapidly as k increases. When $\delta = 0$ and $k = 5$, the power of the Dif-Sar test is remarkably low although the bias is sizable.

The C tests appear to be more powerful than the Dif-Sar test for every case except when $\delta = 2$ and $k = 10$. When the test is modified to be nonnegative, it gains nominal power. Note, however, that the test is oversized. When $k = 40$, the level moment conditions are essentially valid, but the relative rejection frequencies are more than the nominal size of 5%. The test statistics $C(\hat{\alpha}_{d1})$ and $C(\hat{\alpha}_{s1})$ have greater size distortions. To compare the power appropriately by adjusting size, I compare the size-power plots, following Davidson and MacKinnon (1998). Actual size and power are measured as relative rejection frequencies under the null, $\delta = 1$, and under the alternative, $\delta \neq 1$, respectively. As shown in Figure 1, even after size is adjusted, $C(\hat{\alpha}_{d1})$ remains more powerful when $\delta = 0$ and $k = 10$. However, $C(\hat{\alpha}_{d1})$ is the least powerful when $\delta = 2$ and $k = 10$.

The results that the tests perform very differently from one another may be due to the fact that one-step DIF-GMM estimators are not estimated accurately because of the weak instruments problem. The second column in Table 1 shows the DIF-GMM estimators actually are biased downward. Different initial conditions have impacts on the behavior of the DIF-GMM estimators (Hayakawa, 2009). It results in the differences of the test statistics.

When α is smaller so that the series is not persistent and the weak instruments problem is not serious, there is little difference across the tests. Table 2 shows the simulation results in the case of $\alpha = 0.4$. Note that the series converges to its mean-stationary level very quickly

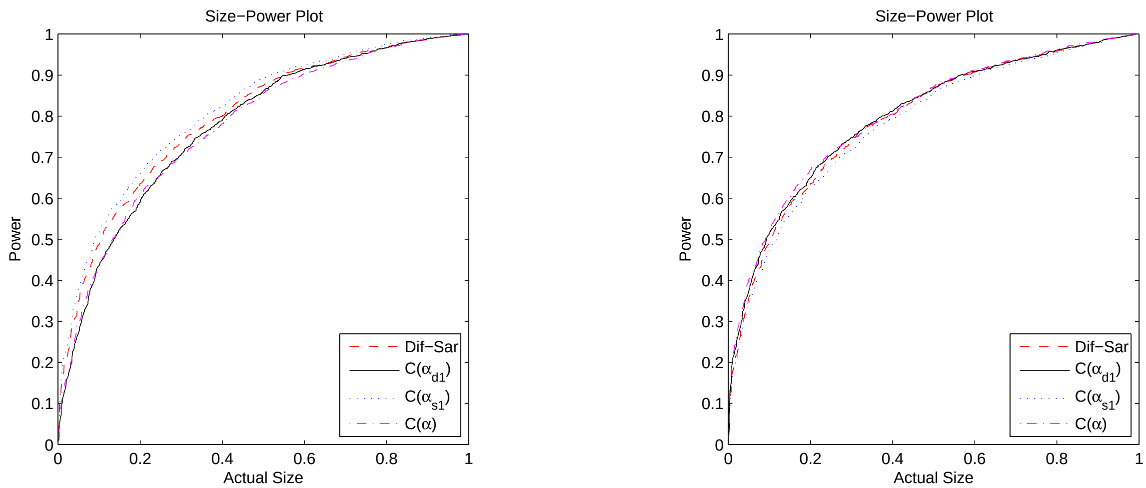
Table 2: Relative Frequencies of Rejection of the Level Moment Conditions: $\alpha = 0.4$

	$\hat{\alpha}_{s2}$	$\hat{\alpha}_{d1}$	Dif-Sar	$C(\hat{\alpha}_d)$	$C(\hat{\alpha}_s)$	$C(\alpha)$
$\delta = 0$						
$k = 0$	0.6222 (0.0498)	0.3824 (0.0538)	1.000	1.000	1.000	1.000
2	0.4381 (0.0468)	0.3816 (0.0574)	0.396	0.314	0.442	0.356
5	0.4063 (0.0406)	0.3848 (0.0525)	0.075	0.076	0.082	0.077
$\delta = 2$						
$k = 0$	0.4970 (0.0335)	0.3934 (0.0319)	1.000	1.000	1.000	1.000
2	0.3996 (0.0404)	0.3861 (0.0495)	0.398	0.415	0.400	0.439
5	0.4038 (0.0401)	0.3851 (0.0520)	0.072	0.074	0.081	0.081

Note: See Table 1. The figures for $k = 10, 20, 40$ are suppressed here since they are almost the same as those for $k = 5$.

Figure 2: The Size-Power Plots: $\alpha = 0.4$

(a) $\delta = 0$ and $k = 2$ (b) $\delta = 2$ and $k = 2$



since α is small. When $k = 5$, there is essentially no bias in $\hat{\alpha}_{s2}$. Even though there seems to be no bias in $\hat{\alpha}_{s2}$, the level moment conditions are rejected by all the tests rather frequently when $k = 2$ for the initial condition $\delta = 2$. Figure 2 shows the size-plots. The curves of all the tests overlap. Unlike the results in the case where $\alpha = 0.8$, the way of calculating the test statistics does not yield significant differences. In the light of these simulation results, we should be careful about the inference on the validity of the level moment conditions, especially when the series is persistent and the weak instruments problem is severe.

5 Application to Data

As a last exercise, I investigate how different test statistics lead to different conclusions, using U.K. manufacturing company data that Arellano and Bond (1991) and Blundell and Bond

Table 3: The Test Statistics of the Overidentifying Restrictions Tests

	$\text{Sar}_d(\hat{\alpha}_{d1})$ (1)	$\text{Sar}_d(\hat{\alpha}_{s1})$ (2)	$\text{Sar}_s(\hat{\alpha}_{d1})$ (3)	$\text{Sar}_s(\hat{\alpha}_{s1})$ (4)	Dif-Sar =(4)-(1)	$C(\hat{\alpha}_{d1})$ =(3)-(1)	$C(\hat{\alpha}_{s1})$ =(4)-(2)
The Sample Period: 1976 -1984							
Test value	88.80	87.46	117.82	112.92	24.13	29.03	25.46
d.f.	(79)	(79)	(100)	(100)	(21)	(21)	(21)
p-value	[0.21]	[0.24]	[0.11]	[0.18]	[0.29]	[0.11]	[0.23]
The Sample Period: 1979 - 1984							
Test value	13.96	19.81	22.25	39.89	25.93	8.29	20.08
d.f.	(25)	(25)	(37)	(37)	(12)	(12)	(12)
p-value	[0.96]	[0.76]	[0.97]	[0.34]	[0.01]	[0.76]	[0.07]

Note: Degrees of freedom and p-values are in round brackets and square brackets, respectively.

(1998) use.[6] They consider the estimation of a dynamic labor demand equation. Following Blundell and Bond (1998), the estimation equation is:

$$n_{i,t} = \alpha_1 n_{i,t-1} + \alpha_2 w_{i,t} + \alpha_3 w_{i,t-1} + \alpha_4 k_{i,t} + \alpha_5 k_{i,t-1} + \lambda_t + \eta_i + \varepsilon_{i,t},$$

where $n_{i,t}$ is the log of employment in firm i in year t, $w_{i,t}$ is the log of the wage rate, and $k_{i,t}$ is the log of the capital stock. Blundell and Bond (1998) treat wages and capital as endogenous variables, and $\Delta w_{i,t-1}$ and $\Delta k_{i,t-1}$ as well as $\Delta n_{i,t-1}$ are used as the instruments in the level equations. Therefore, there are $3 \times (T-2)$ level moment conditions in total. They consider one sample from 1976 to 1984 and its subsample from 1979 to 1984.

Table 3 shows the results. It only reports the test statistics and omits the estimates of the coefficients to save space.[7] In the sample of 1976 to 1984, none of the tests reject the validity of the level moment conditions at conventional levels of significance even though $C(\hat{\alpha}_{d1})$ is only marginally insignificant. However, the tests show much difference results in the sample of 1979 to 1984. Based on Dif-Sar and $C(\hat{\alpha}_{s1})$, the validity of the level moment conditions can be rejected while it $C(\hat{\alpha}_{d1})$ does not reject it. Asymptotic theory does not provide guidance on which test statistic we can rely on.

6 Conclusion

Since the SYS-GMM estimator exploits the level moment conditions in addition to the moment conditions of the DIF-GMM estimator, there is a way to test these specific conditions. Conventionally, the test statistic is calculated as the difference between the Sargan tests of the SYS-GMM and DIF-GMM estimators, where the covariance matrices of the moments are separately estimated. In finite samples, this test statistic can be negative. A relatively straightforward modification can make the test statistic nonnegative.

[6] The data are taken from Stata's online data archive.

[7] Note that only the DIF-GMM estimator for the sample 1976 - 1984 is replicated exactly. The reason their SYS-GMM estimate is not exactly the same as mine is because Blundell and Bond (1998) use a different weighting matrix in one-step SYS-GMM estimation. I use the weighting matrix that is now commonly used. The reason for the difference in results for the 1979-1984 sample is not clear; Stata's built-in command and my MATLAB program produce an identical result that differs from the result of Blundell and Bond (1998).

The simulation results show that when modified, the test gains power, but not in all circumstances. The application to real data shows that the test calculated in different ways actually results in different rejection decisions. Based on this finding, I suggest that in application, the test statistic on the level moment conditions be calculated in the modified ways as well as the conventional way in order to check whether all the ways lead to the same decisions. In future research, I will explore sources of the differences in these forms with a more rigorous approach.

References

Arellano, M. and S. Bond (1991). Some tests of specification for panel data: Monte Carlo evidence and an application to employment equations. *The Review of Economic Studies* ***58***(2), 277–297.

Arellano, M. and O. Bover (1995). Another look at the instrumental variable estimation of error-components models. *Journal of Econometrics* ***68***(1), 29–51.

Blundell, R. and S. Bond (1998). Initial conditions and moment restrictions in dynamic panel data models. *Journal of Econometrics* ***87***(1), 115–143.

Blundell, R. and S. Bond (2000). GMM estimation with persistent panel data: an application to production functions. *Econometric Reviews* ***19***(3), 321–340.

Bowsher, C. G. (2002). On testing overidentifying restrictions in dynamic panel data models. *Economics Letters* ***77***(2), 211–220.

Dahlberg, M., E. Mork, and P. Tovmo (2008). Power properties of the Sargan test in the presence of measurement errors in dynamic panels. *Applied Economics Letters* ***15***(5), 349–353.

Davidson, R. and J. G. MacKinnon (1998). Graphical methods for investigating the size and power of hypothesis tests. *The Manchester School* ***66***(1), 1–26.

Hayakawa, K. (2009). On the effect of mean-nonstationarity in dynamic panel data models. *Journal of Econometrics* ***153***(2), 133–135.

Hayashi, F. (2000). *Econometrics*. Princeton University Press.

Roodman, D. (2009). A note on the theme of too many instruments. *Oxford Bulletin of Economics and Statistics* ***71***(1), 135–158.