# Power Properties of the Sargan Test in the Presence of Measurement Errors in Dynamic Panels‡

Matz Dahlberg#, Eva Mörk♣ and Per Tovmo*

**Abstract**

This paper investigates the power properties of the Sargan test in the presence of measurement errors in dynamic panel data models. The conclusion from Monte Carlo simulations, and an application on the data used by Arellano and Bond (1991), is that in the very likely case of measurement errors in either the dependent or any of the independent variables, we will, if we rely on the Sargan test, quite likely accept a mis-specified model and end up with biased results.

**Keywords**: Sargan test, Measurement errors, Dynamic panels

**JEL Classification**: C12, C15, C23

---

‡ We thank Seung Ahn, Steven Bond, Kåre Johansen, Frank Windmeijer and seminar participants at Uppsala University for helpful comments.

# Corresponding author: Department of Economics, Uppsala University, PO Box 513, SE-751 20 Uppsala, Sweden. E-mail: Matz.Dahlberg@nek.uu.se

♣ The Institute for Labour Market Policy Evaluation, and Department of Economics, Uppsala University.

* Department of Economics, Norwegian University of Science and Technology, Trondheim.

## 1. Introduction

In empirical work, some specification test to test the initial model is desirable. In GMM estimations of dynamic panel data models, the Sargan test for over-identifying restrictions has become the standard one to use. This performs a joint test of the model specification and the validity of the instruments (testing if the moments are fulfilled). However, very little is known about the test's power.[1]

The purpose of this paper is to investigate the power properties of the Sargan test in dynamic panels. Can we feel comfortable with our model specification if the Sargan test does not reject, or might there still be some misspecification leading to serious misinterpretations of the empirical results? In order to answer this question we will perform Monte Carlo simulations where we impose measurement errors in the data, either in the dependent variable or in an independent variable. We will also deliberately impose measurement errors in real data and investigate the consequences for the specification tests and for the estimated coefficients (using the Arellano and Bond (1991) data).

## 2. Measurement errors in $x$

### 2.1 Experimental Design

We start by investigating a case with measurement errors in the independent variable $x$. We use the following data generating process (DGP):

$$y_{it} = \alpha_t + \beta y_{it-1} + \delta x_{it} + f_i + \varepsilon_{it} \tag{1}$$

$$y_{i0} = \frac{f_i}{(1-\beta)} + \upsilon_i \tag{2}$$

[1] An exception to this is Bowsher (2002), where the power properties of the Sargan test is explored when the error term follows an AR(1) process.

where cross-sections are denoted by $i = 1,\ldots,N$ and time periods by $t = 1,\ldots,T$. $\alpha_t$ are time dummies and $f_i$ are individual specific effects. In addition, we let $x$ follow an AR(1)-process:

$$x_{it} = \gamma x_{it-1} + u_{it} \tag{3}$$

In the simulations, we will use sample sizes of $N = (100,\ldots,1000)$ and $T = 7$, and let data be generated by $\beta = 0.5$, $\delta = 1$, $\gamma = (0.5, 0.8)$, $x_{i0} \sim NID(0,1)$, $\varepsilon_{it} \sim NID(0,1)$, $\upsilon_i \sim NID\left(0, 1/\left(1-\beta_1^2\right)\right)$, $u_{it} \sim NID(0,1)$, $f_i \sim NID(0,1)$, and $\alpha_t \sim NID(0,1)$.

To investigate how the Sargan test works when we have problems with measurement errors, we consider three different types of errors.[2] The first one is an additive error, where the observed $x$ (denoted $\hat{x}$) is generated as $\hat{x}_{it} = x_{it} + \eta_{it}$, with the measurement error generated as $\eta_{it} \sim NID(0,1)$ (yielding rather severe measurement errors: the standard deviation of the errors is the same as the standard deviation of *x*) or as $\eta_{it} \sim NID(0, 0.1)$ (yielding less severe errors: the standard deviation of the measurement errors is 10 times smaller than the standard deviation of *x*). The second measurement error is a multiplicative one, where the observed $x$ is generated as $\hat{x}_{it} = x_{it} * \eta_{it}$, with $\eta$ generated as $\eta_{it} \sim NID(1, 0.5)$ and $\eta_{it} \sim NID(1, 0.1)$ respectively. Finally, we consider an exponential measurement error given by $\hat{x}_{it} = x_{it} * e^{\eta_{it}}$, where $\eta_{it} \sim NID(0, 0.1)$.

[2] In Dahlberg, Johansson (now Mörk) and Tovmo (2002), we show that the moments are not fulfilled when there is measurement errors in *x* or *y*.

We estimate equation (1) in first-differenced form, using the GMM-estimator described in Holtz-Eakin, Newey and Rosen (1988) and Arellano and Bond (1991).[3] All Monte Carlo experiments are carried out in GAUSS, using the program's pseudo-random number generator. In each experiment we carry out 1000 Monte Carlo replications.

**2.2 Results**

The results from the simulations with measurement errors in $x$ and when $x$ is treated as exogenous are presented in Table 1. The presented results are for $\gamma = 0.8$[4]. Since we treat $x$ as exogenous, we use contemporaneous $x$ as an instrument (in first-differenced form) as well as lags of $y$ dated two periods back and more.[5]

In the base case, where $x$ is measured without errors, we note from the first panel in Table 1 that the Sargan test has good size properties and that there is virtually no bias in $\delta$. For small sample sizes, there is however a bias in the coefficient for the lagged dependent variable.

When we impose an additive and severe measurement error in $x$, it turns out that the Sargan test has low power at small sample sizes (see the second panel of Table 1). To get a rejection rate over 80 percent when testing at the five percent level, we need a sample of more than 600 cross-sections (shown in Dahlberg, Johansson (now Mörk) and Tovmo (2000)). In applications we would hence far too often accept the false null of a well-specified model when we have a small number of cross-sectional

[3] The weighting matrix we use is the one proposed by Holtz-Eakin, Newey and Rosen.

[4] For $\gamma = 0.5$ and measurement errors in $x$, the Sargan test get worse power properties compared to the $\gamma = 0.8$ case. The results for $\gamma = 0.5$ are available upon request.

[5] When we treat $x$ as endogenous, we use lags of $x$ as instruments (dated one period back and more) instead of contemporaneous values of $x$. The results when $x$ is treated as endogenous, presented in

units. This wouldn't be so problematic if the estimates were unbiased, but here we have a rather severe bias in both $\beta$ and $\delta$.

**Table 1**. Measurement errors in *x*. *x* treated as exogenous.

| | Sargan rejection rates (%) | | | Bias (%) | |
|---|---|---|---|---|---|
| N | 10 | 5 | 1 | $\beta$ | $\delta$ |
| | | | No measurement errors | | |
| 100 | 11.5 | 5.5 | 0.3 | -28.0 | -3.8 |
| 500 | 13.5 | 7.0 | 1.8 | -4.7 | -0.7 |
| 1000 | 10.1 | 5.2 | 1.3 | -2.0 | -0.3 |
| | | | $\hat{x} = x + \eta$, $\eta \sim NID(0,1)$ | | |
| 100 | 28.8 | 13.7 | 4.0 | -8.1 | -64.9 |
| 500 | 80.1 | 70.8 | 47.9 | 43.4 | -62.3 |
| 1000 | 97.8 | 96.3 | 89.1 | 50.8 | -61.8 |
| | | | $\hat{x} = x + \eta$, $\eta \sim NID(0,0.1)$ | | |
| 100 | 12.0 | 5.3 | 0.4 | -27.2 | -5.5 |
| 500 | 13.7 | 6.2 | 1.8 | -3.2 | -2.3 |
| 1000 | 10.9 | 5.0 | 1.3 | -0.5 | -1.8 |
| | | | $\hat{x} = x * \eta$, $\eta \sim NID(1,0.1)$ | | |
| 100 | 13.0 | 5.7 | 0.5 | -26.4 | -7.4 |
| 500 | 14.4 | 7.0 | 1.8 | -1.5 | -4.2 |
| 1000 | 11.3 | 5.9 | 1.6 | 1.3 | -3.8 |
| | | | $\hat{x} = x * \eta$, $\eta \sim NID(1,0.5)$ | | |
| 100 | 21.7 | 12.7 | 2.5 | -12.7 | -50.5 |
| 500 | 63.7 | 51.5 | 28.0 | 31.6 | -47.8 |
| 1000 | 91.2 | 85.1 | 63.6 | 37.6 | -47.3 |
| | | | $\hat{x} = x * e^{\eta}$, $\eta \sim NID(0,0.1)$ | | |
| 100 | 13.0 | 5.6 | 0.5 | -26.4 | -7.9 |
| 500 | 14.7 | 7.0 | 1.7 | -1.5 | -4.7 |
| 1000 | 11.3 | 6.0 | 1.6 | 1.3 | -4.3 |

Note: As instruments we use lags of *y* dated two periods back and more and *x* in first-differenced form. $\beta$ is the coefficient for the lagged dependent variable *y* and $\delta$ is the coefficient for *x*.

In the third panel we investigate the performance when the additive measurement error is less severe. We find that the Sargan test has very low power, but the bias in $\beta$ and $\delta$ is rather low, even though there might be a problem for small sample sizes: for $N = 100$, the bias is approximately 27 percent for $\beta$ and 5 percent for $\delta$. These results are very similar to the cases of multiplicative and exponential measurement errors with a small variation in the errors (c.f. the fourth and last panels).

---

Table 2 in Dahlberg, Johansson (now Mörk) and Tovmo (2002), show a similar pattern as the one found

However, also for multiplicative measurement errors, the results are sensitive to how severe the errors are. When the standard deviation of the measurement errors is half that of *x*, we see from the fifth panel that the Sargan test has bad power when *N* is low and that we have a severe bias in both $\beta$ and $\delta$.

## 3. Measurement errors in *y*

### 3.1 Experimental Design

Let us now consider measurement errors in the dependent variable $y$. Since we estimate a dynamic model, this will induce measurement errors in one of the regressors as well (namely the lagged dependent variable). We will consider three types of measurement errors: one additive ($\hat{y}_{it} = y_{it} + \eta_{it}$), one multiplicative ($\hat{y}_{it} = y_{it} * \eta_{it}$), and one exponential ($\hat{y}_{it} = y_{it} * e^{\eta_{it}}$), where $\eta$ is generated as in Section 2. There is no measurement error in *x* and the rest of the DGP is as above. As instruments we use lags of $y$ dated two periods back or more together with contemporaneous $x$ in first differences.

### 3.2 Results

The Monte Carlo results are given in Table 2. When the error is additive and severe (the second panel) the estimates of $\beta$ are severely biased (60 percent when $N = 1000$) and so are the ones of $\delta$, however to a smaller extent (approximately 8 percent). The power increases as the sample size grows, but is still relatively low: for example, when $N = 500$ the Sargan test rejects the false null in only 40 % of the Monte Carlo simulations (testing at the ten percent significance level). The worst case is when the

---

when *x* is treated as exogenous.

measurement error is exponential (see the last panel). The bias is substantial in both of the estimated coefficients and the power is extremely low; when testing at the ten percent significance level, the Sargan test rejects in only 17 percent of the times even for large *N*.

**Table 2**. Measurement errors in $y$.

| | Sargan rejection rates (%) | | | Bias (%) | |
|---|---|---|---|---|---|
| N | 10 | 5 | 1 | $\beta$ | $\delta$ |
| | | | No measurement errors | | |
| 100 | 11.5 | 5.5 | 0.3 | -28.0 | -3.8 |
| 500 | 13.5 | 7.0 | 1.8 | -4.7 | -0.7 |
| 1000 | 10.1 | 5.2 | 1.3 | -2.0 | -0.3 |
| | | | $\hat{y} = y + \eta$, $\eta \sim NID(0,1)$ | | |
| 100 | 15.3 | 6.7 | 1.4 | -74.8 | -9.3 |
| 500 | 40.4 | 27.5 | 10.6 | -62.2 | -7.6 |
| 1000 | 69.3 | 58.5 | 33.6 | -60.8 | -7.7 |
| | | | $\hat{y} = y + \eta$, $\eta \sim NID(0,0.1)$ | | |
| 100 | 11.6 | 15.7 | 0.3 | -28.8 | -3.9 |
| 500 | 12.3 | 7.1 | 1.6 | -5.3 | -0.8 |
| 1000 | 10.1 | 4.2 | 1.5 | -2.8 | -0.4 |
| | | | $\hat{y} = y * \eta$, $\eta \sim NID(1,0.1)$ | | |
| 100 | 15.0 | 7.1 | 1.0 | -37.0 | -5.6 |
| 500 | 17.1 | 9.4 | 2.8 | -12.2 | -1.8 |
| 1000 | 18.8 | 9.5 | 3.1 | -9.9 | -1.6 |
| | | | $\hat{y} = y * \eta$, $\eta \sim NID(1,0.5)$ | | |
| 100 | 16.8 | 6.9 | 0.8 | -92.6 | -13.4 |
| 500 | 32.1 | 20.6 | 7.5 | -80.0 | -9.8 |
| 1000 | 51.4 | 38.2 | 18.6 | -79.0 | -9.7 |
| | | | $\hat{y} = y * e^{\eta}$, $\eta \sim NID(0,0.1)$ | | |
| 100 | 9.5 | 4.5 | 0.5 | -106.0 | 26.6 |
| 500 | 17.4 | 8.0 | 1.1 | -96.4 | 41.0 |
| 1000 | 17.1 | 9.4 | 2.8 | -95.2 | 42.5 |

Note: As instruments we use lags of $y$ dated two periods back and more and $x$ in first-differenced form. $\beta$ is the coefficient for the lagged dependent variable $y$ and $\delta$ is the coefficient for $x$.

When the error is multiplicative, the size of the error matters for the results. When imposing a small error (see the fourth panel), there is virtually no bias in $\delta$, whereas $\beta$ is biased, even though the bias diminishes when the sample size grows. When we increase the standard deviation in the error (the fifth panel), the bias in $\beta$ increases dramatically even for large *N* (for $N = 1000$, the bias is as large as 80 percent). The power of the Sargan test is however low; when testing at the five-percent level, the Sargan test rejects in only 38 percent of the cases for a sample size as large as 1000.[6]

## 4. Application: The Arellano and Bond (1991) data

As we have seen from the simulations, the Sargan test for overidentifying restrictions often leads us to accept models where the moments are not fulfilled with sometimes severely biased parameter estimates as a result.[7] This is especially true when the number of cross-sectional units is small. But of course, the models in the Monte Carlo experiments are very stylized and parsimonious. Do the results have any bearing in real applications? To investigate this, we will use the data in Arellano and Bond (1991) and re-estimate their employment equations with the difference that we have deliberately imposed measurement errors in the employment and wage variables.[8] Our re-estimations are presented in tables 3 and 4. The first two columns in each of these tables restate the results in Arellano and Bond.

---

[6] The power properties are unaffected of the size of $\gamma$ when there are measurement errors in *y*. However, the lower the autoregressive process in *x* is, the lower is the bias in $\beta$ .

[7] It has been suggested, see for example Bowsher (2002), that the power of the Sargan test can be improved by using fewer moment conditions. Doing this does not solve the problem in our case. Another question is whether the power of the Sargan test can be improved by relying on bootstrap critical values using the GMM bootstrap estimator proposed by Hall and Horowitz (1996). The answer is no. It turns out that in the experiments conducted, the bootstrapped Sargan test almost never rejects a false null.

[8] In their application, they have 140 cross-sectional units (quoted U.K. companies) for the period 1979-1984. The equation we estimate is given by

$$n_{it} = \alpha_1 n_{it-1} + \alpha_2 n_{it-2} + \beta_1 w_{it} + \beta_2 w_{it-1} + \gamma k_{it} + \delta_1 ys_{it} + \delta_2 ys_{it-1} + \lambda_t + \eta_i + \upsilon_{it} \qquad (4)$$

where $n_{it}$ denotes the logarithm of U.K. employment in company *i* at the end of year *t*, $w_{it}$ is the log of the real wage, $k_{it}$ is the log of gross capital, $ys_{it}$ is the log of industry output, $\lambda_t$ is a time effect that is common to all companies, $\eta_i$ is a fixed but unobservable firm-specific effect, and $\upsilon_{it}$ is the error term (for exact definitions of the variables, see Arellano and Bond (1991)). The estimation of equation (4) yields the results in column b in Table 4 in Arellano and Bond (1991).

**Table 3.** Measurement error in wages

| | Arellano and Bond[a] | | Additive error[b] | | Multiplicative error[c] | |
|---|---|---|---|---|---|---|
| **Variable** | **Coeff** | **T-ratio** | **Coeff** | **T-ratio** | **Coeff** | **T-ratio** |
| *n*(-1) | 0.47 | 5.56 | 0.29 | 4.76 | 0.29 | 5.36 |
| *n*(-2) | -0.05 | -1.94 | -0.02 | -1.15 | -0.04 | -1.9 |
| *W* | -0.51 | -10.4 | -0.19 | -3.97 | 0.002 | 0.361 |
| *w*(-1) | 0.22 | 2.81 | -0.004 | -0.08 | -0.01 | -1.77 |
| *K* | 0.29 | 7.42 | 0.32 | 7.65 | 0.32 | 7.47 |
| *Ys* | 0.61 | 5.62 | 0.39 | 3.39 | 0.31 | 2.82 |
| *Ys*(-1) | -0.45 | -3.58 | -0.04 | -0.43 | 0.04 | 0.487 |
| | Statistic | P-value | Statistic | P-value | Statistic | P-value |
| Sargan | 30.11 | (0.220) | 31.35 | (0.178) | 27.17 | (0.347) |
| $m_2$ | -0.33 | (0.739) | -0.00 | (0.998) | 0.18 | (0.858) |

Notes: a) Column b in Table 4 in Arellano and Bond (1991), b) The measurement error is generated as NID(0,1), c) The measurement error is generated as NID(3,1). Time dummies are included in all equations.

In Table 3, we impose measurement errors in one of the independent variables, namely the wage variable. In the middle columns we have imposed additive measurement errors, while we in the final columns have imposed multiplicative measurement errors. In the additive case, the errors are distributed $NID(0,1)$. In the multiplicative case, the errors are distributed $NID(3,1)$. A standard deviation of one corresponds to approximately 18 percent of the standard deviation in the wage variable.[9] Neither the Sargan statistic nor the $m_2$ statistic gives us any reason to believe that something is wrong. As a matter of fact, both the Sargan and the $m_2$ statistics are very similar to the ones obtained when no measurement errors are imposed in the data. However, turning to the parameter estimates, we see that they are indeed affected by the imposed measurement errors. Taking the case of additive errors as an example, we see that the short-run wage elasticity decreases from –0.51 to –0.19 and there seems to be less dynamics when measurement errors are imposed.

9 The standard deviation of the wage variable is 5.6 (unlogged values). We have also estimated with less variation in the measurement errors (we have used distributions of the errors that corresponds to

**Table 4.** Measurement error in employment (the dependent variable)

| | Arellano and Bond[a] | | Additive error[b] | | Multiplicative error[c] | |
|---|---|---|---|---|---|---|
| **Variable** | **Coeff** | **T-ratio** | **Coeff** | **T-ratio** | **Coeff** | **T-ratio** |
| $n(-1)$ | 0.47 | 5.56 | 0.29 | 3.28 | -0.13 | -1.14 |
| $n(-2)$ | -0.05 | -1.94 | 0.10 | 1.56 | -0.10 | -1.87 |
| $w$ | -0.51 | -10.40 | -0.21 | -2.20 | -0.39 | -2.19 |
| $w(-1)$ | 0.22 | 2.81 | 0.01 | 0.06 | -0.22 | -1.36 |
| $Y$ | 0.29 | 7.42 | 0.09 | 1.56 | 0.42 | 5.36 |
| $ys$ | 0.61 | 5.62 | 0.34 | 1.42 | 0.66 | 2.09 |
| $ys(-1)$ | -0.45 | -3.58 | -0.21 | -0.89 | -0.24 | -0.62 |
| | Statistic | P-value | Statistic | P-value | Statistic | P-value |
| Sargan | 30.11 | (0.220) | 32.18 | (0.153) | 28.66 | (0.278) |
| $m_2$ | -0.33 | (0.739) | 0.51 | (0.611) | -0.48 | (0.633) |

Notes: a) Column b in Table 4 in Arellano and Bond (1991), b) The measurement error is generated as NID(5,1), c) The measurement error is generated as NID(5,1). Time dummies are included in all equations.

In Table 4, we impose measurement errors in the dependent variable, i.e. the employment variable. In the middle columns we have imposed additive measurement errors, while we in the final columns have imposed multiplicative measurement errors. In both cases the errors are distributed NID(5,1). A standard deviation of one corresponds to approximately 6.5 percent of the standard deviation in the employment variable.[10] As can be seen from the bottom rows in the table, the Sargan test does not give us any reason to believe that the moments are not fulfilled. Neither does the $m_2$ statistic, which tests for lack of second-order serial correlation in the first-differenced residuals. As for the results in Table 3, both the Sargan and the $m_2$ statistics are very similar to the ones obtained when no measurement errors are imposed in the data. What is affected, though, is the parameter estimates. For example, we seem to end up with less dynamics when there are additive errors: the second lag of both employment and wages is insignificant. The coefficients estimates in the Arellano and Bond

five and ten percent of the distribution in the wage variable). This did, however, not change the results substantially.

10 The standard deviation of the employment variable is 15.9 (unlogged values). The measurement errors are imposed before the variables are logged. The reason for having a mean of five in the errors is to ensure that the resulting employment variable is positive.

estimations suggest a short-run wage elasticity of –0.51 while the corresponding figure with additive errors is –0.21 and with multiplicative errors –0.39.[11]

## 5. Conclusions

The general conclusion from the Monte Carlo simulations and the application on the Arellano and Bond (1991) data is that the Sargan test for overidentifying restrictions often leads us to accept models where the moments are not fulfilled with sometimes severely biased parameter estimates as a result. The problem is most pronounced when the number of cross-sectional units is small and when there are measurement errors in the dependent variable.

Arellano and Bond conclude after their empirical application: "The GMM estimator offers significant efficiency gains compared to simpler IV alternatives, and produces estimates that are well-determined in dynamic panel data models. … The robust $m_2$ statistics perform satisfactorily as do the two-step Sargan …" (p. 293). From the results in this study, we do however think that this statement must be qualified. The Sargan statistic performs satisfactorily and the GMM estimator will produces estimates that are well-determined in dynamic panel data models given that the models are correctly specified. In the very likely case of measurement errors in either the dependent or any of the independent variables, we will with a rather high probability accept a mis-specified model and end up with biased results.

[11] The long-run wage elasticity is –0.5 in the Arellano and Bond estimations, –0.33 with additive errors, and –0.63 with multiplicative errors.

**References**

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, 277-297.

Bowsher, C. (2002), “On Testing Overidentifying Restrictions in Dynamic Panel Data Models”, Economics Letters, 77, 211-220.

Dahlberg, M., E. Johansson, and P. Tovmo (2002), “Power Properties of the Sargan Test in the Presence of Measurement Errors in Dynamic Panels”, Working Paper 2002:13, Department of Economics, Uppsala University.

Hall, P. and J. L. Horowitz (1996), “Bootstrap Critical Values for Tests Based on Generalized-Method-of-Moments Estimators”, *Econometrica* 64, 891-916.

Holtz-Eakin, D., W. Newey, and H. S. Rosen (1988), “Estimating Vector Autoregressions with Panel Data”, *Econometrica* 56, 1371-1395.