SYSTEM GMM ESTIMATION WITH A SMALL SAMPLE

Marcelo Soto*

July 2009

Properties of GMM estimators for panel data, which have become very popular in the empirical economic growth literature, are not well known when the number of individuals is small. This paper analyses through Monte Carlo simulations the properties of various GMM and other estimators when the number of individuals is the one typically available in country growth studies. It is found that, provided that some persistency is present in the series, the system GMM estimator has a lower bias and higher efficiency than all the other estimators analysed, including the standard first-differences GMM estimator.

Keywords: Economic Growth, System GMM estimation, Monte Carlo Simulations
JEL classification: C15, C33, O11

* Institut d'Anàlisi Econòmica, Barcelona. Email: marcelo.soto@iae.csic.es. I am grateful to Richard Blundell, Frank Windmeijer and participants in the Econometric Society meetings in Mexico DF and Wellingtron for comments and helpful suggestions. The support from the Spanish Ministry of Science and Innovation under project ECO2008-04837/ECON is gratefully acknowledged. The author acknowledges the support of the Barcelona GSE Research Network and of the Government of Catalonia.

1. Introduction

The development and application of Generalised Methods of Moments (GMM) estimation for panel data has been extremely fruitful in the last decade. For instance, Arellano and Bond (1991), who pioneered the applied GMM estimation for panel data, have more than 1,200 citations according to ISI Web of Knowledge as of July 2009.

In the empirical growth literature, GMM estimation has become particularly popular. The Arellano and Bond (1991) estimator in particular initially benefited from widespread use in different topics related to growth[1]. Subsequently the related Blundell and Bond (1998) estimator has gained an even grater attention in the empirical growth literature[2].

However, these GMM estimators were designed in the context of labour and industrial studies. In such studies the number of individuals N is large, whereas the typical number of cross-units in economic growth samples is much smaller. Indeed, availability of country data limits N to at most 100 and often to less than half that value.

The lack of knowledge about the properties of GMM estimators when N is small renders them a sort of a black box. Moreover, a practical problem not addressed in the earlier literature refers to fact that the low number of cross-units may prevent the use of the full set of instruments available. This implies that, in order to make estimation possible, the number of instruments must be reduced. The performance of the various GMM estimators in panel data is not well known when only a partial set of instruments is used for estimation.

This paper analyses through Monte Carlo simulations the performance of the system GMM and other standard estimators when the number of individuals is small. The simulations follow closely those made by Blundell et al (2000) in the sense that the structure of the model simulated is exactly the same as theirs. The only difference is that Blundell et al. chose N=500, while this paper reports results for N more adapted to the actual sample size of growth regressions in a panel of countries (N=100, 50, 35). A small N constrains the researcher to limit the number of instruments used for estimation, which may also have a consequence on the properties of the estimators. The paper studies the behaviour of the estimators for different choices on the instruments.

---

1 For instance Caselli et al (1996) use it to test the Solow model; Greeanway et al (2002) for analysing the impact of trade liberalisation in developing countries; and Banerjee and Duflo (2003) to investigate the effect of income inequality on growth.

2 Some examples are studies on aid and growth (Dalgaard et al, 2004); education and growth (Cohen and Soto, 2007); and exchange rate volatility and growth (Aghion et al, 2009).

The next section depicts the econometric model under consideration. Section 3 presents the estimation results obtained by Monte Carlo simulations. Section 4 concludes.

2. The econometric model

We will consider an autoregressive model with one additional regressor:

$$y_{it} = \alpha y_{it-1} + \beta x_{it} + \eta_i + u_{it} \tag{1}$$

for i = 1,…, N and t = 2,…, T, with $|\alpha| < 1$.The disturbances η_i and u_{it} have the standard properties. That is,

$$E(\eta_i) = 0, E(u_{it}) = 0, E(\eta_i u_{it}) = 0 \quad \text{for i = 1,…, N and t = 2,…, T.} \tag{2}$$

Additionally, the time-varying errors are assumed uncorrelated:

$$E(u_{is} u_{it}) = 0 \quad \text{for i = 1,…, N and } \forall \text{ t} \neq \text{s.} \tag{3}$$

Note that no condition is imposed on the variance of u_{it}, hence the moment conditions used below do not require homoskedasticity.
The variable x_{it} is also assumed to follow an autoregressive process:

$$x_{it} = \rho x_{it-1} + \tau\eta_i + \theta u_{it} + e_{it} \tag{4}$$

for i = 1,…, N and t = 2,…, T, with $|\rho| < 1$. The properties of the disturbance e_{it} are analogous to those of u_{it}. More precisely,

$$E(e_{it}) = 0, E(\eta_i e_{it}) = 0 \quad \text{for i = 1,…, N and t = 2,…, T.} \tag{5}$$

Two sources of endogeneity are present in the x_{it} process. First, the fixed-effect component η_i has an effect on x_{it} through a parameter τ –implying that y_{it} and x_{it} have both a steady state determined only by η_i. And second, the time-varying disturbance u_{it} impacts x_{it} with a parameter θ. A situation in which the attenuation bias due to measurement error predominates over the upward bias due to simultaneity determination may be simulated with $\theta < 0$.
For simplicity, it is useful to express x_{it} and y_{it} as deviations from their steady state values. Under the additional hypothesis that (4) is a valid representation of x_{it} for t = 1,…, $-\infty$, x_{it}

may be written as a deviation from its steady state:

$$x_{it} = \frac{\tau\eta_i}{1-\rho} + \xi_{it} \tag{6}$$

where the deviation from steady state ξ_{it} is equal to

$$\xi_{it} \equiv (1-\rho L)^{-1}(\theta u_{it} + e_{it}).$$

In this last expression L is the lag operator and so, for any variable w_{it} and parameter λ, $(1 - \lambda L)^{-1} w_{it}$ is defined as

$$(1-\lambda L)^{-1} w_{it} \equiv w_{it} + \lambda w_{it\text{-}1} + \lambda^2 w_{it\text{-}2} \ldots$$

Similarly, assuming that (1) is a valid representation of y_{it} for $t = 1, \ldots, -\infty$, we have,

$$y_{it} = \frac{\eta_i}{1-\alpha} + \beta(1-\alpha L)^{-1} x_{it} + (1-\alpha L)^{-1} u_{it}$$

After substituting in this last expression x_{it} by (6), y_{it} may be written as

$$y_{it} = \frac{(1-\rho+\beta\tau)}{(1-\alpha)(1-\rho)}\eta_i + \zeta_{it} \tag{7}$$

with its deviation from steady state given by,

$$\zeta_{it} \equiv \beta(1-\alpha L)^{-1}(1-\rho L)^{-1}[\theta u_{it} + e_{it}] + (1-\alpha L)^{-1} u_{it}.$$

Hence, the deviation ζ_{it} from the steady state is the sum of two independent AR(2) processes and one AR(1) process.

3. Monte Carlo simulations

This section reports Monte Carlo simulations for the model described in (1) to (5) and analyses the performance of different estimators. To summarise, the model specification is:

$$y_{it} = \alpha y_{it-1} + \beta x_{it} + \eta_i + u_{it} \tag{8}$$

$$x_{it} = \rho x_{it-1} + \tau\eta_i + \theta u_{it} + e_{it} \quad (9)$$

with

$$\eta_i \sim N(0;\sigma_\eta^2);\ \ u_{it} \sim N(0;\sigma_u^2);\ e_{it} \sim N(0;\sigma_e^2).$$

We will consider three different cases for the autoregressive processes: no persistency ($\alpha = \rho = 0$), moderate persistency ($\alpha = \rho = 0.5$) and high persistency ($\alpha = \rho = 0.95$). The other parameters are kept fixed in each simulation as follows[3]:

$$\beta = 1; \tau = 0.25; \theta = -0.1; \sigma_\eta^2 = 1; \sigma_u^2 = 1; \sigma_e^2 = 0.16.$$

The parameter θ is negative in order to emulate the effects of measurement error in x_{it}[4]. The hypothesis of homoskedasticity is dropped in subsequent simulations. Initially, the sample size considered is N = 100 and T = 5. In later simulations N is set at 50 and 35 with T=12, in order to illustrate the effects of a low number of individuals (relative to T). Each result presented below is based on a different set of 1000 replications, with new initial observations generated for each replication. The appendix A explains with more details the generation of initial observations.

The estimators analysed are OLS, fixed-effects, difference GMM, level GMM and system GMM. One and two-step results are reported for each GMM estimation. All the estimations are performed with the program DPD for Gauss (Arellano and Bond, 1998).

3.1 Accuracy and efficiency results

The main finding is that, provided that some persistency is present in the series, the system GMM estimator yields the results with the lowest bias. Consider Table 1, which presents results for N=100 and T=5. The performance of each estimator varies according to the degree of persistency in the series. For instance, when α and ρ are both equal to zero, OLS estimates wrongly assign a highly significant coefficient to the lagged dependent variable, whereas the Within estimator provides a negative and significant coefficient[5]. However,

---

[3] These values are the same as those selected by Blundell et al (2000)

[4] Hauk and Wacziarg (2009) carry out Monte Carlo simulations for the convergence equation derived from the Solow model by directly introducing noise in the variables.

[5] Recall that the OLS coefficient on y_{it-1} is biased towards 1 and the Within groups coefficient is biased downwards (with a bias decreasing with T).

OLS provides estimates for β with the lowest root mean square error (RMSE) in the no-persistency case[6]. The high RMSE on β displayed by all GMM estimates is a consequence of the weakness of the instruments for x_{it} discussed in appendix B when $\rho = 0$.

In the moderate persistency case (α and ρ equal to 0.5), the OLS estimator has again a strong upwards bias for α and a downwards bias for β. The Within estimator is strongly biased downwards in both cases. The difference GMM estimator results in coefficients between 60% and 70% the real parameter values and presents the highest RMSE for β. This shows that lagged levels are weak instruments for variables in differences even in a moderate-persistency environment. As to the level and system GMM estimates, they display systematically the lowest bias for both α and β. In addition, these estimators result in the lowest RMSE for α.

In the high persistency case (α and ρ equal to 0.95), the system GMM estimator outperform all the other estimators in terms of bias and efficiency. Note however that the bias in the lagged dependent variable of the OLS estimator is considerably reduced. This is due to the fact that this coefficient is biased towards 1.

One well known caveat of GMM estimators refers to their reported two-step standard errors, which systematically underestimate the real standard deviation of the estimates (Blundell et al, 2000). For instance, standard errors of system GMM are 62% to 74% lower than the standard deviation of the estimates of α and 70% to 83% lower in the case of β. This result suggests taking the one-step estimates for inference purposes, since accuracy and efficiency (measured by the RMSE) are similar to those of the two-steps. The variance correction suggested by Windmeijer (2005) is implemented in the current simulations.

The next step is to replicate the Monte Carlo experiments by changing the sample sizes. Results are now obtained by setting N = 50 and T = 12. The reduction of N relative to T precludes the use of the full set of instruments derived from conditions (10) and (10). Indeed, if all those moment conditions were exploited the number of instruments would be $(T-2)(T+1)$, which exceeds N. On the other hand, the optimal weighting matrix W_S defined in (10) has a rank of N at most. Therefore, if the number of instruments exceeds N, W_S is singular and the two-step estimator cannot be computed. In order to make estimation possible only the most relevant (i.e. the most recent) instruments are used in

---

[6] The RMSE on β is defined as $\sqrt{\frac{1}{R}\sum_{j=1}^{R}\left(\beta_j - \beta\right)^2}$ where R is the total number of replications.

each period. That means that only levels lagged two periods are used for the equation in differences and, as before, differences lagged one period are used for the equation in levels. This procedure results in 4(T−2) instruments. The results are presented in Table 2.

The main conclusions are the same as those obtained from Table 1. That is, in the simulations without persistency, all the estimators perform badly. The OLS and fixed-effect estimators present considerable biases and, although the system GMM estimator displays a relatively low bias, it has a high RMSE for β. As to the moderate and high-persistency cases, the system GMM does better than any other estimator overall. Still, the moderate-persistency estimation suffers from a small sample upwards bias of 20% for α and of 14% for β.

The next set of results is obtained by straining even more the sample size, with N = 35 and T = 12. From the previous discussion it becomes apparent that the system GMM estimation with the set of instruments used in the previous simulations is not feasible since the number of instruments 4×(12−2)=40 exceeds the number of individuals. Several alternatives were considered for reducing the number of instruments. First, lagged levels of x_{it} were omitted from the instrument set for the equation in differences and Z_l was kept as before. Second, lagged differences of x_{it} were omitted from the instrument set for the equation in levels and Z_d was kept as before. And third, Z_l was kept as before and Z_d was modified as follows,

$$\mathbf{Z}_{di} = \begin{bmatrix} y_{i1} & & & 0 & x_{i1} \\ & y_{i2} & & & x_{i2} \\ & & \ddots & & \vdots \\ 0 & & & y_{iT-2} & x_{iT-2} \end{bmatrix} \qquad (10)$$

Under this last alternative the total number of instruments is 3×(T−2)+1=31. Although all three alternatives provided similar results in terms of bias, the third alternative resulted in the lowest RMSE error in the high-persistency case. Table 3 compares the different estimators, with the instruments for the system GMM estimator defined in (10) for the equation in differences and in (10) for the equation in levels. In general the RMSE are higher than in the previous simulations due to the smaller sample size. In addition, the upward bias for α obtained by system GMM is higher than before. But overall, this estimator outperforms in terms of accuracy and efficiency once again.

The last case under consideration is when errors are heteroskedastic across individuals. The results presented in tables 1 to 3 are based on residuals with variances $\sigma_u^2 = 1$ and $\sigma_e^2 = 0.16$. Now heteroskedasticity is introduced by generating residuals u_{it} and e_{it} such that $\sigma_{u_i}^2 \sim U(0.5\,; 1.5)$ and $\sigma_{e_i}^2 = 0.16\sigma_{u_i}^2$. This particular structure implies that the expected variances of u_{it} and e_{it} are the same as in the previous simulations and that the ratio $\sigma_{u_i}^2 / \sigma_{e_i}^2$ is constant, thus making easier the comparison with the results previously shown. Table 4 reports the simulation results with N = 35 and T =12. The instruments used correspond to those of Table 3. The main effect of heteroskedasticity is to slightly increase the RMSE of β in the high-persistency case. Still, the level and system GMM estimators display both the lowest finite-sample bias and the lowest RMSE.

Figure 1 presents the distribution of the estimates obtained with OLS, and one and two-step system GMM. The distributions correspond to the sample with N = 35, T = 12 and with heteroskedastic residuals across individuals. The vertical line corresponds to the parameter values. The figure shows that the distributions of the one and two-step system GMM are more concentrated around parameter α than the distribution obtained from OLS. However, the GMM estimators are systematically biased upwards, though the bias is considerably lower than the one present in OLS. Regarding β, the distributions of GMM estimates in the no-persistency case –though centred on the right value– have very fat queues. The OLS estimator performs better in this particular case. In the cases with moderate and high persistency the dispersion of GMM distribution is considerably reduced and its bias is systematically lower than OLS's.

One striking feature of GMM estimators is that the gain in efficiency from the two-step estimator is almost inexistent: the one and two step distributions are virtually the same. More work should be done in order find out the cases in which the two-step estimator does better than the one-step estimator.

3.2 Type-I error and power of significance tests.

Another aspect in which the various estimators can be evaluated is the frequency of wrong rejections of the hypothesis that α is not significant when it is in fact equal to zero (i.e. type-I error) and the power to properly reject lack of significance when coefficients are different from zero. Table 5 reports the frequency rejections at a 5% level of the hypothesis that α = 0 and β = 0 for the different simulations described above.

One striking feature –which came out already from figure 1– is that the OLS estimator always rejects lack of significance, even in the case when $\alpha = 0$ (see the no-persistency case). This is an additional flagrant implication of the upward bias of OLS estimates. A similar phenomenon occurs with the Within estimator, which fails to discard significance of α in 43% to 99.5% of the simulations with $\alpha = 0$. As mentioned before, the standard error of two-step GMM estimators underestimate the real variability of the coefficients. A consequence of this is the relatively high number of wrong rejections of non-significance of α in the simulations with $\alpha = 0$ in two-step GMM estimates (up to 69% in the system GMM estimator in the simulation with heteroskedasticity). The lowest type-I errors correspond to one-step GMM estimators, although they also tend to over-reject as N becomes smaller.

The weakness of the difference GMM estimator is reflected in its low power to reject non-significance when parameters are in fact different from zero. For instance, in the high-persistency case the one-step difference GMM estimator rejects non-significance of β in only 4% to 30% of the simulations –that is, it wrongly dismisses the significance of β in 70% to 96% of the simulations. The system estimator is the most powerful among GMM estimators, with its power increasing as series become more persistent. For instance, according to one-step estimates in the heteroskedastic case, the non-significance of β was rejected in 56% of the simulations without persistency, 92% of simulations with moderate persistency, and 100% of simulations with high persistency.

Overall, the one-step system GMM is the more reliable estimator in terms of power and error type-I. Among all the estimators presented in the table, the OLS estimator has the highest power in absolute terms (it never rejected significance in the simulations). But the counterpart of this is that inference based on OLS estimates is a poor guide when the decision of rejecting a potential non-significant but endogenous variable comes up.

3.3 Overidentifying restrictions tests

One crucial feature of instrumental variables is their exogeneity. Frequency rejections of overidentifying restriction tests –in which the null hypothesis is that instruments are uncorrelated with u_{it}– are presented in Table 6. By construction, the instruments used for estimation are all exogenous, so one would expect that at a 5% level, exogeneity would be rejected in 5% of the simulations. In samples with N = 100 and T = 5 there is a slight tendency towards under-reject exogeneity in the two extreme cases of persistency. But in simulations with smaller N and larger T the under-rejection is much more accentuated. In fact, the system GMM estimation results in overidentifying restriction tests that (properly) never reject exogeneity. However, the fact that the frequency of rejections is lower than their expected value, suggests that small sample bias is affecting the tests. In order to understand better the consequences of this bias, simulations with autocorrelated residuals should be made. When residuals u_{it} are autocorrelated lagged levels or differences of the regressors would be correlated with the u_{it}, hence they could not be used as instruments. This kind of simulations was not performed in this paper.

4. Conclusions

This paper has analysed the properties of recently developed GMM and other standard estimators, obtained with Monte Carlo simulations. Although earlier studies by Blundell and Bond (1998) and Blundell et al (2000) have already shown the superiority of the system GMM estimator over other estimators, the validity of their results when the number of individuals is small are largely unknown. Understanding better the properties of these estimators when N is small is important given the popularity that this method of estimation is getting in empirical growth studies.

A small number of individuals –that is, the number of countries typically available in panel studies– does not seem to have important effects on the properties previously outlined about the system GMM estimator. Namely, when series are moderately or highly persistent, this estimator presents the lowest bias and highest precision. As expected, the OLS estimator has the lowest variance, but the gain in terms of accuracy that the system GMM estimator presents, makes it a more reliable tool in the practical work. Other widely used estimators –the fixed effect and the difference GMM– are systematically outperformed by the system GMM estimator.

Moreover, the properties of this estimator are not hindered when N is so small that it is not possible to exploit the full set of linear moment conditions. This is also an important finding since the earlier Monte Carlo simulations were carried out by using the full set of instruments, whereas in practical work this may not be always feasible, especially in the context of country growth studies.

Overall the system GMM estimator displays the best features in terms of small sample bias and precision. This, together with its simple implementation convert it into a powerful tool in applied econometrics. The next step is to contrast the results provided by this new estimator in applied econometrics with those obtained by the standard estimators.

Appendix A: Construction of initial observations

In order to reduce truncation error, which is may be important particularly in simulations with high persistency, initial observations are built as follows (see Kiviet (1995) for a more general discussion on the generation of initial observations). First, let's omit the index i and rewrite (6) as:

$$x_t = \frac{\tau\eta}{1-\rho} + \xi_t \tag{A.1}$$

where ξ_t represents the deviation of x_t from its steady state and is given by:

$$\begin{aligned} \xi_t &\equiv (1-\rho L)^{-1}(\theta u_t + e_t) \\ &= \theta p_t + q_t \end{aligned} \tag{A.2}$$

So ξ_t is the sum of two independent AR(1) processes, p_t and q_t, with variances

$$\sigma_p^2 = \frac{\sigma_u^2}{1-\rho^2} \text{ and } \sigma_q^2 = \frac{\sigma_e^2}{1-\rho^2} \ .$$

Based on independent random variables η, u_1 and e_1, it is possible to generate initial stationarity-consistent observations x_1 as follows:

$$x_1 = \frac{\tau\eta}{1-\rho} + [\theta u_1 + e_1](1-\rho^2)^{-0.5} \ .$$

Similarly, rewriting equation (7) –where y_t is represented as a deviation from steady state, we obtain:

$$y_t = \frac{(1-\rho+\beta\tau)}{(1-\alpha)(1-\rho)}\eta + \zeta_t . \tag{A.3}$$

The deviation from steady state ζ_t is given by,

$$\zeta_t \equiv \beta(1-\alpha L)^{-1}(1-\rho L)^{-1}[\theta u_t + e_t] + (1-\alpha L)^{-1} u_t . \tag{A.4}$$

According to the expression (A.4) ζ_t is a sum of two independent AR(2) processes and one AR(1) process, and so it can be expressed as:

$$\zeta_t = \beta\theta r_t + \beta s_t + v_t \tag{A.5}$$

where

$r_t = \phi_1 r_{t-1} + \phi_2 r_{t-2} + u_t$,

$s_t = \phi_1 s_{t-1} + \phi_2 s_{t-2} + e_t$ and

$v_t = \alpha v_{t-1} + u_t$

The autoregressive parameters are $\phi_1 = \alpha + \rho$ and $\phi_2 = -\alpha\rho$. The variances of each one of these processes are given by:

$\sigma_r^2 = \sigma_u^2(1-\phi_2)\big/\left[(1+\phi_2)\left((1-\phi_2)^2 - \phi_1^2\right)\right]$

$\sigma_s^2 = \sigma_e^2(1-\phi_2)\big/\left[(1+\phi_2)\left((1-\phi_2)^2 - \phi_1^2\right)\right]$

$\sigma_v^2 = \sigma_u^2\big/\left(1-\alpha^2\right)$

With these variances at hand it is possible to generate for each individual initial observations of r_1, s_1 and v_1 that are consistent with mean and variance stationarity. Then initial observations of y_1 are generated based on equations (A.3) and (A.5) as follows:

$$y_1 = \frac{(1-\rho+\beta\tau)}{(1-\alpha)(1-\rho)}\eta + \beta\theta r_1 + \beta s_1 + v_1$$

Since this procedure does not ensure covariance stationarity of initial observations, each series y_t was generated for 50 + T periods and the first 50 periods were deleted for estimation.

Appendix B: The GMM estimator

This appendix describes the difference and system GMM estimators and the problem of weak instruments. Much of the material of this appendix is based on Arellano and Bond (1991) and Blundell and Bond (1998).

B.1 First-difference GMM estimator

The standard Arellano and Bond (1991) estimator consists in taking equation (1) in first differences and then using y_{it-2} and x_{it-2} for t = 3,…, T as instruments for changes in period t. The exogeneity of these instruments is a consequence of the assumed absence of serial correlation in the disturbances u_{it}. Namely, there are z_d = (T−1)(T−2) moment conditions implied by the model that may be exploited to obtain z_d different instruments. The moment conditions are:

$$E(y_{it-s}\Delta u_{it}) = 0 \text{ and } E(x_{it-s}\Delta u_{it}) = 0 \tag{B.1}$$

for t = 3,…, T and s = 2,…, t−1, where $\Delta u_{it} = u_{it} - u_{it-1}$. Thus for each individual i, restrictions (B.1) may be written compactly as,

$$E(\mathbf{Z}'_{di}\Delta\mathbf{u}_i) = 0 \tag{B.2}$$

where $\mathbf{Z}_{di}$ is the (T − 2)×z_d matrix given by,

$$\mathbf{Z}_{di} = \begin{bmatrix} y_{i1}\ x_{i1} & & & 0 \\ & y_{i1}\ y_{i2}\ x_{i1}\ x_{i2} & & \\ & & \ddots & \\ 0 & & & y_{i1} \cdots y_{iT-2}\ x_{i1} \cdots x_{iT-2} \end{bmatrix} \tag{B.3}$$

and $\mathbf{u}_i$ is the (T − 2)×1 vector $(u_{i3}, u_{i4},\ldots,u_{iT})'$.

Setting the matrix $\mathbf{Z}_d = (\mathbf{Z}'_{d1}, \mathbf{Z}'_{d2},\ldots, \mathbf{Z}'_{dN})'$, the matrix **X** formed by the stacked matrices $\mathbf{X}_i = ((y_{i2}\ x_{i3})', (y_{i3}\ x_{i4})',\ldots, (y_{iT-1}\ x_{iT})')'$ and the vector **Y** formed by the stacked vectors $\mathbf{Y}_i = (y_{i3}, y_{i3},\ldots, y_{iT})'$, the GMM estimation of $\mathbf{B} = (\alpha, \beta)'$ based on the moment conditions (B.2) is given by,

$$\mathbf{B}_d = \left(\Delta\mathbf{X}'\mathbf{Z}_d(\mathbf{W}_d)^{-1}\mathbf{Z}'_d\Delta\mathbf{X}\right)^{-1}\left(\Delta\mathbf{X}'\mathbf{Z}_d(\mathbf{W}_d)^{-1}\mathbf{Z}'_d\Delta\mathbf{Y}\right) \tag{B.4}$$

where $\mathbf{W}_d$ is some $z_d \times z_d$ positive definite matrix. From equation (B.4) it can be seen that the standard instrumental variable estimator with instruments given by (B.3) is a particular case of the GMM estimator. Indeed, the standard IV estimator is obtained by letting $\mathbf{W}_d = \mathbf{Z}'_d\mathbf{Z}_d$. Hansen (1982) shows that the matrix $\mathbf{W}_d$ yielding the optimal (i.e., minimum variance) GMM estimator based only in moment conditions (B.2) is,

$$\mathbf{W}_d = \sum_{i=1}^{N} \mathbf{Z}_{di}^{'} \Delta \mathbf{u}_i \Delta \mathbf{u}_i^{'} \mathbf{Z}_{di} \qquad \text{(B.5)}$$

Since the actual vectors of errors $\Delta\mathbf{u}_i$ are unknown, a first step estimation is needed in order to make Hansen's estimator operational. Although no knowledge is required about the variance of $\mathbf{u}_i$, Arellano and Bond (1991) suggest taking into account the variance structure of the differenced error term $\Delta\mathbf{u}_i$ that would result under the assumption of homoskedasticity. In that case, $E(\Delta\mathbf{u}_i\Delta\mathbf{u}'_i) = \sigma_u^2 \times \mathbf{G}$ where $\mathbf{G}$ is the $(T-2)\times(T-2)$ matrix given by,

$$\mathbf{G} = \begin{bmatrix} 2 & -1 & & & \mathbf{0} \\ -1 & 2 & -1 & & \\ & -1 & & & \\ & & & \ddots & -1 \\ \mathbf{0} & & & -1 & 2 \end{bmatrix}$$

Therefore, the one-step Arellano-Bond estimator is obtained by using,

$$\mathbf{W}_d = \sum_{i=1}^{N} \mathbf{Z}_{di}^{'} \mathbf{G} \mathbf{Z}_{di} \qquad \text{(B.6)}$$

Then, the two-step estimator is obtained by substituting $\Delta\mathbf{u}_i$ in (B.5) by the residuals from the one-step estimation.

B.2 Weak instruments

Blundell and Bond (1998) show that the first-difference GMM estimator for a purely autoregressive model –i.e. without additional regressors– has a large finite sample bias and low precision in two cases. First, when the autoregressive parameter α tends to unity and second, when the variance of the specific-effect η_i increases with respect to the variance of u_{it}. In both cases lagged levels of the dependent variable become weaker instruments since they are less correlated with subsequent changes. To see this clearly, Blundell and Bond consider the simple case of T = 3. In this case only one observation per individual is available for estimation, which leads to the following single instrumental variable equation:

$$\Delta y_{i2} = (\alpha - 1)y_{i1} + \eta_i + u_{i2} \text{ , for i = 1,…, N.} \tag{B.7}$$

The least-squares estimation of $(\alpha - 1)$ in equation (B.7) yields the following coefficient,

$$\hat{\pi} = \frac{\sum_i^N \Delta y_{i2} y_{i1}}{\sum_i^N y_{i1}^2} \text{ ,}$$

which has a probability limit equal to,

$$\text{Plim } \hat{\pi} = (\alpha - 1)\frac{k}{k + \sigma_\eta^2 / \sigma_u^2} \text{ , where } k = \frac{1-\alpha}{1+\alpha}.$$

Not surprisingly the probability limit of $\hat{\pi}$ tends towards zero as y_{it} approaches to a random walk process. Also, due to the positive correlation between y_{i1} and η_i, $\hat{\pi}$ tends towards zero as σ_η^2 increases relative to σ_u^2.

A similar result is obtained when additional regressors are included as in the model (1)-(5). To show this keeping tractability, consider again the simple case with T = 3 and with x_{i1} being the only lagged variable used as an instrument. In that case the single instrumental variable equation for Δx_{i3} is

$$\Delta x_{i3} = \rho(\rho - 1)x_{i1} + \rho\tau\eta_i + \upsilon_{i3} \tag{B.8}$$

for i = 1,…, N and where $\upsilon_{i3} = \theta u_{i3} + e_{i3} + (\rho - 1)(\theta u_{i2} + e_{i2})$. The least square estimate for $\rho(\rho - 1)$ is given by,

$$\hat{\pi} = \frac{\sum_i^N \Delta x_{i3} x_{i1}}{\sum_i^N x_{i1}^2},$$

which has a probability limit equal to,

$$\text{Plim } \hat{\pi} = \frac{\rho(\rho-1)}{1+\gamma}, \text{ where } \gamma = \frac{\tau^2\sigma_\eta^2}{\theta^2\sigma_u^2 + \sigma_e^2}.$$

As in the Blundell-Bond example of a purely autoregressive equation, lagged levels of an additional regressor are weak instruments as ρ tends to 1 and when the variance of the specific effect component of x_{i1}, $\tau^2\sigma_\eta^2$, is large relative to the variance of its temporal disturbance, $\theta^2\sigma_u^2 + \sigma_e^2$. The only difference with the purely autoregressive case is that when the autoregressive parameter ρ tends to zero, x_{i1} is also a weak instrument for Δx_{i3}. This is the result of using levels of variables lagged two periods as instruments for contemporary changes.

As Staiger and Stock (1997) show, when instruments are weakly correlated with the regressors, instrumental variable methods have a strong bias and the standard errors underestimate the real variability of the estimators. Therefore, in the context of growth regressions where –even considering a time trend– variables like the income level or human and physical capital stocks have a strong autoregressive component, the standard difference GMM estimator is not likely to perform satisfactorily. Hence the need in growth regressions of GMM estimators that deal with the problem of high persistency in the series.

B.3 System GMM estimator

Arellano and Bover (1995) and Blundell and Bond (1998) suggest to use lagged differences as instruments for estimating equations in levels. The validity of these instruments require only a mild condition on initial values, which is

$$E[(u_{i3} + \eta_i)\Delta x_{i2}] = 0 \text{ and } E[(u_{i3} + \eta_i)\Delta y_{i2}] = 0 \tag{B.9}$$

Conditions (B.9) together with the model set out in (1) to (5) imply the following moment conditions:

$$E[(u_{it} + \eta_i)\Delta x_{it-1}] = 0 \text{ and } E[(u_{it} + \eta_i)\Delta y_{it-1}] = 0 \tag{B.10}$$

for i = 1,…, N and t = 3,…,T. So, if conditions (B.9) are true, (B.10) results in the existence of $z_l = 2(T-2)$ instruments for the equation in levels.

In fact conditions (B.9) are always true under the hypothesis of mean stationarity implicit in the derivation of expressions (6) and (7) and the absence of serial correlation in u_{it}. Indeed, under such hypothesis we have:

$$E[(u_{i3} + \eta_i)\Delta x_{i2}] = E[(u_{i3} + \eta_i)\Delta \xi_{i2}] = 0$$

and

$$E[(u_{i3} + \eta_i)\Delta y_{i2}] = E[(u_{i3} + \eta_i)\Delta \zeta_{i2}] = 0.$$

The advantage of estimation in levels is that lagged differences are informative about current levels of variables even when the autoregressive coefficients approach unity. To see this, consider again the simple case of T = 3 and where Δx_{i2} is the only variable used as instrument for x_{i3}. Then the first-stage instrumental variable equation is

$$x_{i3} = \rho\Delta x_{i2} + (1+\rho)\tau\eta_i + \upsilon_{i3}$$

where now υ_{i3} is a function of u_{it} and e_{it} for t = 3, 1 and earlier dates. The least square estimator for ρ has a probability limit equal to

$$\text{Plim } \hat{\pi} = \frac{\rho}{2}.$$

Therefore, lagged differences remain informative about current levels even when the autoregressive parameter tends to one. On the other hand, when the autoregressive parameter ρ tends to 0, lagged differences lose their explanatory power.

Blundell and Bond (1998) propose to exploit conditions (B.10) by estimating a system of equations formed by the equation in first-differences and the equation in levels. The instruments used for the equation in first-differences are those described above, while the instruments for the equation in levels for each individual i are given by the $(T-2)\times z_l$ matrix

$$\mathbf{Z}_{li} = \begin{bmatrix} \Delta y_{i2}\ \Delta x_{i2} & & & 0 \\ & \Delta y_{i3}\ \Delta x_{i3} & & \\ & & \ddots & \\ 0 & & & \Delta y_{iT-1}\ \Delta x_{iT-1} \end{bmatrix} \tag{B.11}$$

Thus, letting the matrix $\mathbf{Z}_l = (\mathbf{Z}'_{l1}, \mathbf{Z}'_{l2}, \ldots, \mathbf{Z}'_{lN})'$, the whole set of instruments used in the system GMM estimator is given by the $2N(T-2)\times(z_d + z_l)$ matrix

$$\mathbf{Z}_s = \begin{bmatrix} \mathbf{Z}_d & 0 \\ 0 & \mathbf{Z}_l \end{bmatrix}$$

The one-step system GMM estimator is obtained with the weighting matrix $\mathbf{W}_S$ defined as:

$$\mathbf{W}_s = \begin{bmatrix} \mathbf{W}_d & 0 \\ 0 & \mathbf{W}_l \end{bmatrix},$$

(B.12)

where $\mathbf{W}_l = \sum_{i=1}^{N} \mathbf{Z}_{li}^{'}\mathbf{Z}_{li}$ and $\mathbf{W}$d is defined in (B.6). So the one-step estimator is:

$$\mathbf{B}_S = \left(\mathbf{X}_S^{'}\mathbf{Z}_S(\mathbf{W}_S)^{-1}\mathbf{Z}_S^{'}\mathbf{X}_S\right)^{-1}\left(\mathbf{X}_S^{'}\mathbf{Z}_S(\mathbf{W}_S)^{-1}\mathbf{Z}_S^{'}\mathbf{Y}_S\right)$$

where $\mathbf{X}_s$ is a stacked matrix of regressors in differences and levels and where $\mathbf{Y}_s$ is a stacked vector of the dependent variable in differences and levels. Finally, the heteroskedasticity-robust two-step estimator is obtained as explained above.

It is straightforward to show that the system GMM estimator is a weighted average of the difference and level coefficients. Indeed, defining the matrices,

$$\mathbf{P}_m = \mathbf{Z}_m\mathbf{W}_m^{-1}\mathbf{Z}_m^{'} \text{ and } \mathbf{Q}_m = \mathbf{X}_m^{'}\mathbf{P}_m\mathbf{X}_m$$

where m = d, l or s (i.e., matrices with variables in differences, levels or system), and noting that

$$\mathbf{Q}_S = \mathbf{Q}_d + \mathbf{Q}_l$$

the system estimator $\mathbf{B}_s$ may be written as:

$$\mathbf{B}_S = \mathbf{Q}_S^{-1}\left(\mathbf{X}_d^{'}\mathbf{P}_d\mathbf{Y}_d + \mathbf{X}_l^{'}\mathbf{P}_l\mathbf{Y}_l\right)$$

$$= \mathbf{Q}_S^{-1}\mathbf{Q}_d\mathbf{B}_d + \left(\mathbf{I} - \mathbf{Q}_S^{-1}\mathbf{Q}_d\right)\mathbf{B}_l$$

In this last expression B_d and B_l are the difference and level estimators, respectively, and **I** is the identity matrix. The weight on B_d is:

$$\mathbf{Q}_S^{-1}\mathbf{Q}_d = \left[\pi_d^{'}\mathbf{Z}_d^{'}\mathbf{Z}_d\pi_d + \pi_l^{'}\mathbf{Z}_l^{'}\mathbf{Z}_l\pi_l\right]^{-1}\left[\pi_d^{'}\mathbf{Z}_d^{'}\mathbf{Z}_d\pi_d\right]$$

where π_d is a $z_d \times 2$ matrix of coefficients obtained by least square in the underlying first-stage regression of $\mathbf{X}_d$ on $\mathbf{Z}_d$, and π_l is defined analogously. Therefore, as the explanatory power of the instruments for the equation in differences decreases and π_d tends towards zero, the system GMM estimator tends towards the level GMM estimator.

References

Aghion, P., Bacchetta, P., Ranciere, R., Rogoff, K., 2009. Exchange rate volatility and productivity growth: The role of financial development. Journal of Monetary Economics 56(4), 494-513.

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

Arellano, M., Bond, S., 1998. Dynamic panel data estimation using DPD98 for Gauss.

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

Banerjee, A.., Duflo, E., 2003. Inequality and growth: What can the data say? Journal of Economic Growth, 8, 267-299.

Blundell, R., Bond, S., 1998. Initial conditions and moment restrictions in dynamic panel data models. Journal of Econometrics 87, 115-143.

Blundell, R., Bond, S., Windmeijer, F., 2000. Estimation in dynamic panel data models: Improving on the performance of the standard GMM estimator, in Baltagi, B. (ed.), Nonstationary Panels, Panel Cointegration, and Dynamic Panels. Advances in Econometrics 15. JAI Press, Elsevier Science, Amsterdam.

Caselli F., Esquivel, G., Lefort, F., 1996. Reopening the convergence debate: a new look at cross-country growth empirics. Journal of Economic Growth 1, 363-389.

Cohen, D., Soto, M., 2007. Growth And Human Capital: Good Data, Good Results. Journal of Economic Growth 12(1), 51-76.

Dalgaard, C., Hansen, H., Tarp F., 2004. On the empirics of foreign aid and growth. Economic Journal 114(496), F191-F216.

Greenaway, D., Morgan, W., Wright, P., 2002. Trade liberalisation and growth in developing countries. Journal of Development economics 67(1), 229-244.

Hansen, L., 1982. Large sample properties of generalized method of moments estimators. Econometrica 50, 1029-1054.

Hauk, W., Wacziarg, R., 2009. A Monte Carlo study of growth regressions. Journal of Economic Growth 14, 103-147.

Kiviet, J., 1995. On bias, inconsistency and efficiency of various estimators in dynamic panel data models. Journal of Economietrics 68, 53-78.

Staiger, D., Stock, H., 1997. Instrumental variables regression with weak instruments. Econometrica 65(3), 557-586.

Windmeijer, F., 2005. A finite sample correction for the variance of linear efficient two-step GMM estimators. Journal of Econometrics 126(1), 25-51.

Table 1: Simulations with N = 100 and T = 5

		α				$\beta = 1$			
	Estimator	Mean	RMSE	Std. Dev.	Std. Err. / Std. Dev.	Mean	RMSE	Std Dev	Std. Err. / Std. Dev.
	OLS	0.493	0.496	0.045	1.002	0.977	0.137	0.135	0.988
	Within	-0.242	0.247	0.050	0.965	0.394	0.621	0.136	0.961
	DIF GMM - 1	-0.027	0.100	0.096	1.024	0.395	1.018	0.819	0.980
No	DIF GMM - 2	-0.027	0.107	0.103	0.868	0.390	1.059	0.865	0.848
Persistency	LEV GMM - 1	0.038	0.118	0.112	1.103	1.572	1.780	1.685	1.059
($\alpha = \rho = 0$)	LEV GMM - 2	0.029	0.121	0.117	1.001	1.544	1.832	1.750	0.968
	SYS GMM - 1	0.019	0.089	0.087	1.004	0.886	0.786	0.778	1.014
	SYS GMM - 2	0.021	0.089	0.086	0.740	0.844	0.811	0.796	0.694
	OLS	0.820	0.321	0.022	0.980	0.773	0.249	0.103	0.982
	Within	0.136	0.368	0.055	1.000	0.388	0.630	0.146	0.995
	DIF GMM - 1	0.368	0.212	0.166	0.993	0.653	0.633	0.529	0.969
Moderate	DIF GMM - 2	0.363	0.227	0.181	0.829	0.632	0.686	0.579	0.805
Persistency	LEV GMM - 1	0.577	0.133	0.109	1.017	1.174	0.581	0.555	1.000
($\alpha = \rho = 0.5$)	LEV GMM - 2	0.566	0.135	0.118	0.894	1.165	0.606	0.583	0.910
	SYS GMM - 1	0.552	0.113	0.100	0.966	1.067	0.413	0.408	0.981
	SYS GMM - 2	0.556	0.117	0.103	0.653	1.032	0.416	0.414	0.722
	OLS	0.963	0.013	0.002	1.010	0.886	0.124	0.049	1.014
	Within	0.749	0.206	0.041	0.994	0.574	0.453	0.154	0.980
	DIF GMM - 1	0.895	0.100	0.084	0.993	0.285	1.208	0.974	1.012
High	DIF GMM - 2	0.891	0.109	0.092	0.831	0.254	1.283	1.044	0.860
Persistency	LEV GMM - 1	0.958	0.011	0.007	1.102	0.991	0.127	0.127	1.127
($\alpha = \rho = 0.95$)	LEV GMM - 2	0.958	0.011	0.008	1.012	0.988	0.133	0.132	1.035
	SYS GMM - 1	0.958	0.011	0.007	1.087	0.990	0.113	0.113	1.158
	SYS GMM - 2	0.958	0.011	0.008	0.625	1.002	0.111	0.111	0.828

Table 2: Simulations with N = 50 and T = 12

		α				$\beta = 1$			
	Estimator	Mean	RMSE	Std. Dev.	Std. Err. / Std. Dev.	Mean	RMSE	Std Dev	Std. Err. / Std. Dev.
	OLS	0.493	0.495	0.047	0.956	0.968	0.122	0.118	0.987
	Within	-0.084	0.094	0.042	0.972	0.406	0.602	0.100	1.028
	DIF GMM - 1	-0.025	0.065	0.060	0.987	0.400	0.723	0.403	0.962
No	DIF GMM - 2	-0.020	0.067	0.064	0.418	0.399	0.738	0.428	0.418
Persistency	LEV GMM - 1	0.074	0.105	0.074	0.984	1.565	0.920	0.727	0.949
($\alpha = \rho = 0$)	LEV GMM - 2	0.065	0.105	0.082	0.646	1.508	0.967	0.822	0.609
	SYS GMM - 1	0.041	0.077	0.065	0.982	1.103	0.546	0.536	0.954
	SYS GMM - 2	0.043	0.082	0.069	0.364	1.074	0.545	0.540	0.289
	OLS	0.818	0.319	0.019	0.967	0.781	0.235	0.086	0.980
	Within	0.393	0.114	0.039	0.990	0.530	0.480	0.101	1.015
	DIF GMM - 1	0.410	0.131	0.095	0.896	0.704	0.410	0.285	0.970
Moderate	DIF GMM - 2	0.411	0.136	0.103	0.368	0.702	0.431	0.312	0.415
Persistency	LEV GMM - 1	0.638	0.153	0.066	0.993	1.205	0.378	0.318	0.989
($\alpha = \rho = 0.5$)	LEV GMM - 2	0.631	0.151	0.076	0.638	1.195	0.403	0.352	0.672
	SYS GMM - 1	0.601	0.120	0.064	0.979	1.140	0.319	0.287	1.008
	SYS GMM - 2	0.601	0.122	0.069	0.338	1.126	0.325	0.300	0.357
	OLS	0.963	0.013	0.002	0.950	0.883	0.126	0.046	0.947
	Within	0.922	0.032	0.016	0.960	0.815	0.203	0.084	0.959
	DIF GMM - 1	0.913	0.054	0.040	0.932	0.475	0.692	0.451	0.974
High	DIF GMM - 2	0.910	0.059	0.044	0.377	0.457	0.736	0.497	0.387
Persistency	LEV GMM - 1	0.959	0.009	0.004	1.074	0.988	0.083	0.082	1.062
($\alpha = \rho = 0.95$)	LEV GMM - 2	0.959	0.010	0.004	0.714	0.983	0.093	0.092	0.710
	SYS GMM - 1	0.958	0.009	0.004	1.055	0.990	0.083	0.083	1.037
	SYS GMM - 2	0.958	0.009	0.004	0.373	0.993	0.089	0.088	0.380

Table 3: Simulations with N = 35 and T = 12

		α				$\beta = 1$			
	Estimator	Mean	RMSE	Std. Dev.	Std. Err. / Std. Dev.	Mean	RMSE	Std Dev	Std. Err. / Std. Dev.
	OLS	0.495	0.498	0.057	0.943	0.966	0.152	0.148	0.929
	Within	-0.087	0.101	0.051	0.956	0.401	0.612	0.126	0.967
	DIF GMM - 1	-0.028	0.084	0.079	0.977	0.387	0.841	0.576	0.945
No	DIF GMM - 2	-0.023	0.091	0.088	0.546	0.398	0.893	0.660	0.522
Persistency	LEV GMM - 1	0.100	0.134	0.089	0.954	1.510	0.842	0.669	0.997
($\alpha = \rho = 0$)	LEV GMM - 2	0.094	0.136	0.098	0.537	1.466	0.872	0.737	0.554
	SYS GMM - 1	0.060	0.101	0.081	0.954	1.231	0.616	0.571	0.982
	SYS GMM - 2	0.060	0.104	0.084	0.291	1.190	0.601	0.571	0.250
	OLS	0.818	0.319	0.023	0.930	0.774	0.250	0.107	0.938
	Within	0.391	0.119	0.047	0.974	0.529	0.488	0.126	0.958
	DIF GMM - 1	0.423	0.139	0.116	0.945	0.723	0.487	0.400	0.969
Moderate	DIF GMM - 2	0.422	0.150	0.128	0.522	0.715	0.519	0.434	0.555
Persistency	LEV GMM - 1	0.661	0.177	0.074	0.966	1.155	0.383	0.350	0.975
($\alpha = \rho = 0.5$)	LEV GMM - 2	0.656	0.177	0.084	0.539	1.149	0.407	0.378	0.576
	SYS GMM - 1	0.624	0.145	0.074	0.948	1.148	0.367	0.336	0.966
	SYS GMM - 2	0.623	0.145	0.076	0.277	1.127	0.365	0.342	0.294
	OLS	0.963	0.013	0.002	0.935	0.887	0.126	0.055	0.932
	Within	0.920	0.035	0.019	0.936	0.817	0.211	0.105	0.904
	DIF GMM - 1	0.921	0.056	0.048	0.987	0.619	0.671	0.553	0.999
High	DIF GMM - 2	0.917	0.067	0.058	0.475	0.603	0.745	0.631	0.472
Persistency	LEV GMM - 1	0.959	0.010	0.005	1.026	0.986	0.099	0.098	1.032
($\alpha = \rho = 0.95$)	LEV GMM - 2	0.959	0.010	0.005	0.582	0.983	0.110	0.109	0.585
	SYS GMM - 1	0.958	0.010	0.005	1.008	0.990	0.098	0.097	1.011
	SYS GMM - 2	0.959	0.010	0.005	0.324	0.987	0.102	0.101	0.332

Table 4: Simulations with N = 35 and T = 12 and heteroskedasticity across individuals

		α				$\beta = 1$			
	Estimator	Mean	RMSE	Std. Dev.	Std. Err. / Std. Dev.	Mean	RMSE	Std Dev	Std. Err. / Std. Dev.
	OLS	0.492	0.495	0.059	0.938	0.971	0.149	0.146	0.984
	Within	-0.089	0.103	0.053	0.962	0.404	0.610	0.133	0.962
	DIF GMM - 1	-0.033	0.090	0.083	0.961	0.401	0.822	0.564	0.969
No	DIF GMM - 2	-0.027	0.094	0.090	0.512	0.394	0.878	0.635	0.502
Persistency	LEV GMM - 1	0.098	0.135	0.094	0.935	1.425	0.821	0.702	0.959
($\alpha = \rho = 0$)	LEV GMM - 2	0.090	0.135	0.101	0.491	1.398	0.879	0.784	0.493
	SYS GMM - 1	0.056	0.103	0.086	0.933	1.185	0.637	0.610	0.936
	SYS GMM - 2	0.056	0.104	0.088	0.249	1.158	0.632	0.612	0.214
	OLS	0.818	0.319	0.023	0.988	0.776	0.248	0.106	0.969
	Within	0.394	0.116	0.048	0.975	0.525	0.491	0.126	0.994
	DIF GMM - 1	0.429	0.136	0.116	0.965	0.708	0.504	0.411	0.958
Moderate	DIF GMM - 2	0.429	0.145	0.126	0.514	0.708	0.540	0.454	0.510
Persistency	LEV GMM - 1	0.664	0.181	0.075	0.980	1.146	0.386	0.358	0.977
($\alpha = \rho = 0.5$)	LEV GMM - 2	0.660	0.180	0.083	0.517	1.152	0.412	0.383	0.538
	SYS GMM - 1	0.629	0.149	0.075	0.958	1.134	0.361	0.335	0.991
	SYS GMM - 2	0.627	0.149	0.077	0.254	1.117	0.359	0.340	0.273
	OLS	0.963	0.013	0.002	0.925	0.882	0.131	0.057	0.926
	Within	0.920	0.036	0.020	0.932	0.812	0.215	0.104	0.946
	DIF GMM - 1	0.920	0.060	0.052	0.947	0.592	0.716	0.588	0.961
High	DIF GMM - 2	0.915	0.071	0.062	0.444	0.585	0.760	0.637	0.466
Persistency	LEV GMM - 1	0.959	0.010	0.005	0.950	0.983	0.109	0.108	0.991
($\alpha = \rho = 0.95$)	LEV GMM - 2	0.959	0.011	0.005	0.529	0.980	0.116	0.115	0.545
	SYS GMM - 1	0.959	0.010	0.005	0.953	0.985	0.106	0.105	0.989
	SYS GMM - 2	0.959	0.010	0.005	0.277	0.983	0.110	0.109	0.289

Table 5: Frequency rejections of the null hypothesis that coefficients are not significant (at a 5% level)

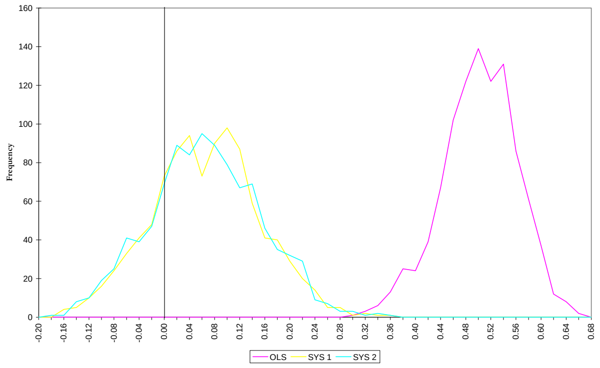
		N=100; T=5		N=50; T=12		N=35; T=12		N=35; T=12; Heteroskedasticity	
	Estimator	α	β	α	β	α	β	α	β
	OLS	1000	1000	1000	1000	1000	1000	1000	1000
	Within	995	853	538	975	456	895	426	870
	DIF GMM - 1	63	64	80	202	93	130	98	122
No	DIF GMM - 2	109	117	448	599	316	397	351	398
Persistency	LEV GMM - 1	59	208	189	643	247	652	219	586
($\alpha = \rho = 0$)	LEV GMM - 2	74	235	332	769	469	819	497	792
	SYS GMM - 1	65	233	111	587	142	623	139	559
	SYS GMM - 2	164	412	531	927	632	960	687	946
	OLS	1000	1000	1000	1000	1000	1000	1000	1000
	Within	714	764	1000	997	1000	986	1000	987
	DIF GMM - 1	651	280	975	733	939	499	933	482
Moderate	DIF GMM - 2	695	333	998	914	980	732	985	731
Persistency	LEV GMM - 1	972	632	1000	964	1000	901	1000	907
($\alpha = \rho = 0.5$)	LEV GMM - 2	970	636	1000	981	1000	967	1000	973
	SYS GMM - 1	994	769	1000	971	1000	924	1000	921
	SYS GMM - 2	999	862	1000	999	1000	994	1000	995
	OLS	1000	1000	1000	1000	1000	1000	1000	1000
	Within	1000	962	1000	1000	1000	1000	1000	1000
	DIF GMM - 1	999	43	1000	281	1000	308	1000	306
High	DIF GMM - 2	999	97	1000	613	1000	601	1000	576
Persistency	LEV GMM - 1	1000	990	1000	1000	1000	1000	1000	1000
($\alpha = \rho = 0.95$)	LEV GMM - 2	1000	990	1000	1000	1000	1000	1000	1000
	SYS GMM - 1	1000	995	1000	1000	1000	1000	1000	1000
	SYS GMM - 2	1000	997	1000	1000	1000	1000	1000	1000

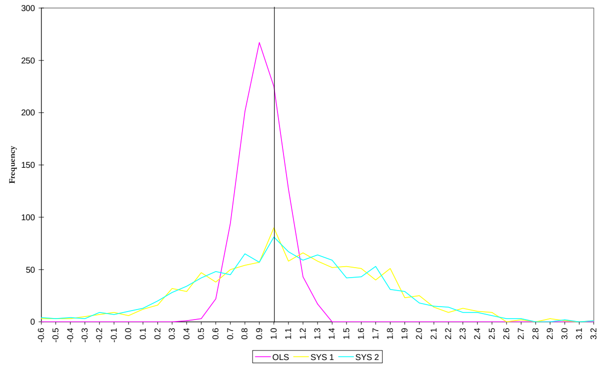
Table 6: Frequency rejections of the null hypothesis that instruments are exogenous (Sargan test)

	Estimator	N=100; T=5	N=50; T=12	N=35; T=12	N=35; T=12; Heterosk.
No	DIF GMM	24	0	1	2
Persistency	LEV GMM	39	9	0	0
$(\alpha = \rho = 0)$	SYS GMM	29	0	0	0
Moderate	DIF GMM	54	0	1	1
Persistency	LEV GMM	74	31	2	0
$(\alpha = \rho = 0.5)$	SYS GMM	49	0	0	0
High	DIF GMM	21	0	1	1
Persistency	LEV GMM	25	10	1	0
$(\alpha = \rho = 0.95)$	SYS GMM	34	0	0	0

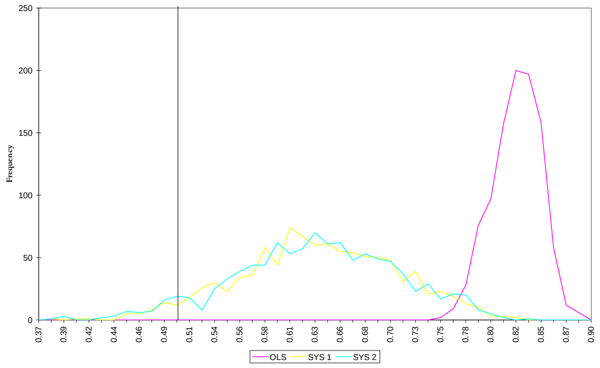
Figure 1: Distributions of estimates in model with N=35, T=12 and heteroskedasticity across individuals

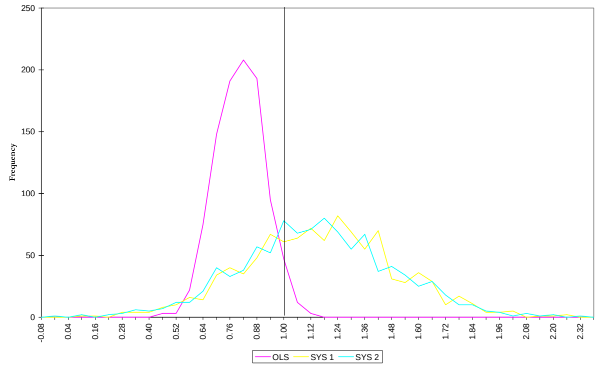
No persistency ($\alpha = 0$; $\beta = 1$)





Moderate persistency ($\alpha = 0.5$; $\beta = 1$)





High persistency ($\alpha = 0.95$; $\beta = 1$)