Journal of Econometrics 234 (2023) 101–110

Contents lists available at ScienceDirect

Journal of Econometrics

journal homepage: www.elsevier.com/locate/jeconom

Initial conditions and Blundell–Bond estimators[☆]

Richard Blundell [a,*], Steve Bond [b]

[a] University College London and Institute for Fiscal Studies, UK
[b] Nuffield College and Department of Economics, University of Oxford, UK

ARTICLE INFO

Article history:
Received 25 January 2023
Received in revised form 25 January 2023
Accepted 25 January 2023
Available online 3 February 2023

JEL classification:
C01
C23
C36
C52

Keywords:
Panel data
Fixed effects
Persistent series
Weak instruments
System estimator

ABSTRACT

We place the Blundell–Bond paper in the context of the early development of panel data estimators that accounted for unobserved heterogeneity, dynamics and persistent economic series. The initial work focused on appropriate econometric methods to estimate dynamic models using unbalanced panel data with many firms and/or individuals but covering a small number of time periods. Eliminating the unobserved firm-specific 'fixed' effects by taking first-differences and using as instruments suitably lagged values of the dependent variable, and of endogenous or predetermined explanatory variables, led to the first-differenced GMM estimators popularised by Arellano and Bond (1991). This approach was less well suited to models which relate highly persistent series. Following Arellano and Bover (1995) we examined the use of suitably lagged first-differences as instruments for the equations in levels and derived the conditions, particularly on initial conditions, under which first-differences of the dependent variable would or would not be uncorrelated with individual-specific 'fixed' effects. An influential contribution was to illustrate the magnitude of the bias when the first-differenced GMM estimator is used to estimate autoregressive models for highly persistent series, and the potential to reduce that bias by using additional valid moment conditions for the equations in levels — thereby popularising the use of these extended or 'System' GMM estimators.

1. Introduction

By the late 1980s, new panel data sources on firms and on households were rapidly becoming available and were accompanied by a flurry of substantive questions on the drivers of productivity, investment, innovation, labour demand, labour supply, and earnings. These were pressing questions in the economics research and policy community. They brought together researchers in microeconometrics and in empirical economics with a shared sense of excitement and collaboration. Reliable answers to these questions required an empirical methodology that incorporated dynamics in behaviour and accounted for the time-series properties of the underlying economic series. The likely importance of differences across people and firms, underlying concerns about unobserved heterogeneity, further motivated the use of panel data.

The origins of the Blundell and Bond (1998) paper date back to this period. More precisely to 1987 and a research programme at the Institute for Fiscal Studies (IFS), funded by the UK Economic and Social Research Council (ESRC), which

☆ We thank all our incredible colleagues who were engaged in pushing forward the dynamic panel data research agenda in the late 1980s and 1990s. Blundell would like to thank the ESRC Centre for the Microeconomic Analysis of Public Policy (CPP) at the Institute for Fiscal Studies (ES/T014334/1) for support.
* Corresponding author.
E-mail address: r.blundell@ucl.ac.uk (R. Blundell).

https://doi.org/10.1016/j.jeconom.2023.01.020
0304-4076/

aimed to develop microeconometric models of company investment using panel data from annual company accounts. Blundell, who had recently been appointed as Research Director at IFS, approached Manuel Arellano for advice on appropriate econometric methods to estimate dynamic models using unbalanced panel data for many firms but covering a small number of time periods. Arellano, who was then a postdoc at the University of Oxford, advised that unobserved firm-specific 'fixed' effects could be eliminated by taking first-differences, with suitably lagged values of the dependent variable, and of endogenous or predetermined explanatory variables, available to use as instruments in the resulting system of first-differenced equations, provided that the time-varying component of the error term of the equations in levels was serially uncorrelated. Lars Hansen's (1982) Generalised Method of Moments (GMM) provided a natural framework to obtain asymptotically efficient estimators based on these moment conditions.

Arellano's advice was followed and implemented in Blundell et al. (1992), one of the first papers to estimate the Hayashi (1982) Q model of investment using firm-level panel data.[1] Spotting the wider potential of this approach, Blundell encouraged Arellano to collaborate with Bond, who was then a pre-doctoral researcher working on the investment project at IFS, both to investigate thoroughly the properties of the first-differenced GMM estimator, and to develop software which would make the estimator more accessible to other researchers. This collaboration led to the widely cited Arellano and Bond (1991) paper, and to the development of their DPD code for Gauss, which undoubtedly contributed to the success of that paper.

Notwithstanding the success of the first-differenced GMM estimator in estimating models which related stationary series like company investment rates and measures of average Q, and the desirable properties illustrated by the Monte Carlo analysis in Arellano and Bond (1991), doubts were expressed about the reliance on lagged levels of variables as instruments for subsequent first-differences. Such concerns were raised most memorably in discussions with Zvi Griliches, whose chief interest was in the potential to use the method to estimate firm-level production functions, relating logarithms of firm-level input and output measures. These production variables exhibit time-series properties very close to random walks, with little correlation between first-differences and lagged levels, suggesting that the instruments used by the first-differenced GMM estimator could be worryingly weak.

The use of additional moment conditions for equations in levels was discussed in Arellano and Bond (1991), but only in the context of explanatory variables which themselves are assumed to be uncorrelated with the time-invariant component of the error term. The breakthrough in Arellano and Bover (1995) was the observation that if the covariance between an explanatory variable and the time-invariant error component is not zero but is constant over time, then first-differences of that explanatory variable are uncorrelated with that error component, and suitably lagged first-differences are available as instruments for the equations in levels, in addition to the moment conditions for the first-differenced equations discussed in Arellano and Bond (1991).

This idea was quickly incorporated into the DPD software, and started to be used in empirical work. Our initial plan for what became the Blundell and Bond (1998) paper was simply to write a survey of those recent developments, covering both Griliches' concern about the weakness of the instruments used by the first-differenced GMM estimator for models relating highly persistent series, and the potential advantages of using additional moment conditions for the equations in levels, with suitably lagged differences as the instruments. The key importance of initial conditions in designing improved estimators for dynamic models was becoming more clearly understood, for example in Blundell and Smith (1991). However, as we began to work on that survey, it became apparent that there were too many unanswered questions simply to review the existing contributions. While there was already a general literature characterising the consequences of using weak instruments, we found no formal discussion in the setting of first-differenced equations for autoregressive panel data models with lagged levels used as instruments. While the Monte Carlo design in Arellano and Bover (1995) used an AR(1) specification, we found no discussion of the conditions under which first-differences of the *dependent variable* would or would not be uncorrelated with individual-specific 'fixed' effects in that setting. And how did using these additional linear moment conditions for the equations in levels compare to using the additional non-linear moment conditions which had been mentioned briefly by Arellano and Bond (1991) and discussed extensively by Ahn and Schmidt (1995)?

Consequently our focus shifted towards filling those gaps. Although the main impact of our paper may well have stemmed from the Monte Carlo results, which illustrated both the magnitude of the bias when the first-differenced GMM estimator is used to estimate autoregressive models for highly persistent series – as suspected by Griliches – and the potential for reducing that bias by using additional valid moment conditions for the equations in levels – thereby popularising the use of these extended or 'system' GMM estimators – we also managed to include sufficient analytical content for the paper to be published in the *Journal of Econometrics*.

2. Some econometric history

Here we briefly provide the dynamic panel data setting for the Blundell–Bond estimator.

[1] Allowing for an AR(1) component in the error term of the Q model was found to be important for obtaining valid moment conditions. This was a specification that we returned to in the Blundell and Bond (2000) paper on the estimation of production functions.

2.1. Dynamic linear models

Consider

$$y_{it} = \alpha y_{i,t-1} + x_{it}\beta + (\eta_i + v_{it}) \qquad |\alpha| < 1$$

for $i = 1, \ldots, N$ and $t = 2, \ldots, T$. Note the initial observation is y_{i1}, so that the first available equation is

$$y_{i2} = \alpha y_{i1} + x_{i2}\beta + (\eta_i + v_{i2})$$

and we have $T - 1$ equations in levels.[2] First suppose $E[x_{it}\eta_i] = 0$ and $E[x_{it}v_{it}] = 0$. Two important properties of the lagged dependent variable are immediately evident: $E[y_{i,t-1}\eta_i] > 0$ since η_i is part of the process that generates $y_{i,t-1}$ according to our specification. Similarly $E[y_{i,t-1}v_{i,t-1}] > 0$. Thus $y_{i,t-1}$ is correlated with the individual effects, and is not strictly exogenous.

Focusing on the simpler dynamic model

$$y_{it} = \alpha y_{i,t-1} + (\eta_i + v_{it}) \qquad |\alpha| < 1$$

it is useful to briefly establish the properties of pooled OLS and Within Groups estimators in this setting. Assuming $E[y_{i,t-1}v_{it}] = 0$, then $p\lim \widehat{\alpha}_{OLS} > \alpha$ as a result of the positive correlation between $y_{i,t-1}$ and η_i. Now consider the within-transformed explanatory variable

$$\widetilde{y}_{i,t-1} = y_{i,t-1} - \frac{1}{T-1}(y_{i1} + \cdots + y_{i,T-1})$$

and error term

$$\widetilde{v}_{it} = v_{it} - \frac{1}{T-1}(v_{i2} + \cdots + v_{iT})$$

Notice that all correlations of order $\frac{1}{T-1}$ are negative, e.g. $corr(y_{i,t-1}, \frac{-1}{T-1}v_{i,t-1})$ and $corr(v_{it}, \frac{-1}{T-1}y_{it})$. This suggests that $E[\widetilde{y}_{i,t-1}\widetilde{v}_{it}] < 0$ and is of order $\frac{1}{T-1}$ (i.e. $E[\widetilde{y}_{i,t-1}\widetilde{v}_{it}] \to 0$ as $(T-1) \to \infty$). These properties can be shown more formally, e.g. Nickell (1981). Thus $p\lim_{N\to\infty} \widehat{\alpha}_{WG} < \alpha$ for fixed T.

The inconsistency of the Within Groups estimator is of order $\frac{1}{T-1}$ and the bias can be substantial when $T < 10$, and can remain non-negligible when $T < 20$. Also note that the inconsistency does not disappear as $\alpha \to 0$. So, unless T is large, the Within estimator does not provide reliable evidence on whether a lagged dependent variable should be included in the model or not. However, in practice it is useful to know that, for the coefficient on a lagged dependent variable, OLS levels is likely to be biased upwards, and (in short panels) Within Groups is likely to be biased downwards. Supposedly consistent estimators that give $\widehat{\alpha} \gg \widehat{\alpha}_{OLS}$ or $\widehat{\alpha} \ll \widehat{\alpha}_{WG}$ should be viewed with suspicion.

2.2. Instrumental variables

A popular class of estimators that are consistent as $N \to \infty$ with T fixed first transform the model to eliminate the individual effects, and then apply instrumental variables.

The Within transformation is not so useful in this context, since it introduces the shocks from all time periods into the transformed error term, $\widetilde{v}_{it}$. That is, a valid instrument z_{it} must satisfy the strict exogeneity assumption $E[z_{it}v_{is}] = 0$ for all s, t. The first-differencing transformation is more promising. Taking first-differences, we have

$$\Delta y_{it} = \alpha \Delta y_{i,t-1} + \Delta v_{it}$$

for $i = 1, \ldots, N$ and $t = 3, \ldots, T$.

First-differenced OLS is not consistent (as $N \to \infty$ or $T \to \infty$, or both), since $\Delta y_{i,t-1} = y_{i,t-1} - y_{i,t-2}$ and $\Delta v_{it} = v_{it} - v_{i,t-1}$, we have

$$E[\Delta y_{i,t-1}\Delta v_{it}] < 0$$

even if $E[y_{i,t-1}v_{it}] = 0$. However, if we are willing to assume that $E[y_{i,t-1}v_{it}] = 0$, then $y_{i,t-2}$ or $\Delta y_{i,t-2}$ are valid instrumental variables for $\Delta y_{i,t-1}$ in the first-differenced equations. Two-stage least squares (2SLS) estimators of this type were suggested by Anderson and Hsiao (1981) e.g.

$$\widehat{\alpha}_{AH} = (\Delta y'_{-1}Z(Z'Z)^{-1}Z'\Delta y_{-1})^{-1}\Delta y'_{-1}Z(Z'Z)^{-1}Z'\Delta y$$

where Δy is the stacked $N(T-2)\times 1$ vector of observations on Δy_{it}, Δy_{-1} is the stacked $N(T-2)\times 1$ vector of observations on $\Delta y_{i,t-1}$, and Z is the stacked $N(T-2) \times 1$ vector of observations on $y_{i,t-2}$. One further time series observation is lost if $\Delta y_{i,t-2}$ rather than $y_{i,t-2}$ is used as the instrument.

[2] Some authors assume y_{i0} is observed, and thus have T equations in levels.

The assumption that $y_{i,t-1}$ is predetermined follows naturally from assuming that the v_{it} are serially uncorrelated shocks, provided the initial conditions (y_{i1}) are also uncorrelated with subsequent v_{it} shocks. Note that a minimum of $T = 3$ time series observations are required to identify α using this approach. With $T = 3$, we have one instrument y_{i1} for Δy_{i2} in the equation

$$\Delta y_{i3} = \alpha \Delta y_{i2} + \Delta v_{i3} \quad \text{for } i = 1, \ldots, N$$

The Anderson–Hsiao 2SLS estimators are consistent as $N \to \infty$ for fixed T. But they are not efficient, except in the special case with $T = 3$. With $T > 3$, further valid instruments become available for the first-differenced equations in the later time periods. Efficiency can be improved by exploiting these additional instruments.

The transformed error term Δv_{it} also has a known moving average form of serial correlation, under the maintained assumption that v_{it} is serially uncorrelated. More generally, Δv_{it} may be heteroskedastic. These features can be exploited to improve efficiency when $T > 3$ (i.e. when the parameter α is overidentified).

2.3. Generalised method of moments (GMM)

GMM formulates a set of orthogonality restrictions (moment conditions) related to an econometric model, and finds parameter estimates that come as close as possible to achieving these orthogonality properties in the sample. Holtz-Eakin et al. (1988) and Arellano and Bond (1991) applied the generalised method of moments approach developed by Hansen (1982) to exploit this additional information in the dynamic panel data problem.

For the AR(1) panel data model we assume the following:

Assumption (error components)

$$E(\eta_i) = E(v_{it}) = E(\eta_i v_{it}) = 0 \;\text{ for } t = 2, \ldots, T$$

Assumption (serially uncorrelated shocks)

$$E(v_{is} v_{it}) = 0 \;\text{ for } s \neq t$$

Assumption (predetermined initial conditions)

$$E(y_{i1} v_{it}) = 0 \;\text{ for } t = 2, \ldots, T$$

These assumptions specify a finite number of linear moment conditions, which can be exploited using a linear GMM estimator.

First-differenced equations	Valid instruments
$(y_{i3} - y_{i2}) = \alpha(y_{i2} - y_{i1}) + (v_{i3} - v_{i2})$	y_{i1}
$(y_{i4} - y_{i3}) = \alpha(y_{i3} - y_{i2}) + (v_{i4} - v_{i3})$	y_{i1}, y_{i2}
$\vdots$	
$(y_{iT} - y_{i,T-1}) = \alpha(y_{i,T-1} - y_{i,T-2}) + (v_{iT} - v_{i,T-1})$	$y_{i1}, y_{i2}, \ldots, y_{i,T-2}$

Clearly $E(y_{i1}\Delta v_{i3}) = 0$ follows from assuming predetermined initial conditions, and $E(y_{i1}\Delta v_{i4}) = 0$ follows analogously. Since $y_{i2} = \alpha y_{i1} + \eta_i + v_{i2}$, the stated assumptions imply $E(y_{i2}\Delta v_{i4}) = 0$.

Similar arguments establish the $m = (T-2)(T-1)/2$ moment conditions

$$E(y_{i,t-s}\Delta v_{it}) = 0 \;\text{ for } t = 3, \ldots, T \text{ and } s \geq 2$$

This gives the set of valid instruments proposed in the previous table. These can also be written as $E(Z_i' \Delta v_i) = 0$ where

$$Z_i = \underset{(T-2)\times m}{\begin{pmatrix} y_{i1} & 0 & 0 & \ldots & 0 & 0 & \ldots & 0 \\ 0 & y_{i1} & y_{i2} & \ldots & 0 & 0 & \ldots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & & \vdots \\ 0 & 0 & 0 & \ldots & y_{i1} & y_{i2} & \ldots & y_{i,T-2} \end{pmatrix}} \text{ and } \Delta v_i = \underset{(T-2)\times 1}{\begin{pmatrix} \Delta v_{i3} \\ \Delta v_{i4} \\ \vdots \\ \Delta v_{iT} \end{pmatrix}}$$

and we define the sample analogue

$$b_N(\alpha) = \frac{1}{N}\sum_{i=1}^{N} Z_i' \Delta v_i(\alpha)$$

For $T = 3$, we have 1 moment condition $E(y_{i1}\Delta v_{i3}) = 0$ and 1 parameter. α is just identified, the choice of the weight matrix is irrelevant, and the optimal GMM estimator coincides with the Anderson–Hsiao 2SLS estimator (using the level $y_{i,t-2}$ as the instrument). For $T > 3$, we have $m > 1$ moment conditions. α is overidentified.

GMM estimators minimise a weighted quadratic distance

$$\begin{aligned}\widehat{\alpha}_{GMM} &= \arg\min_{\alpha} J_N(\alpha) = b_N(\alpha)' W_N b_N(\alpha)\\ &= \arg\min_{\alpha}\left(\frac{1}{N}\sum_{i=1}^{N}\Delta v_i' Z_i\right) W_N \left(\frac{1}{N}\sum_{i=1}^{N} Z_i' \Delta v_i\right)\\ &= (\Delta y_{-1}' Z W_N Z' \Delta y_{-1})^{-1} \Delta y_{-1}' Z W_N Z' \Delta y\end{aligned}$$

for some positive definite m *x* m weight matrix W_N, where Δy and Δy_{-1} are the stacked $N(T-2)\times 1$ vectors of observations on Δy_{it} and $\Delta y_{i,t-1}$ as before, and $Z = (Z_1, \ldots, Z_N)'$ is the stacked $N(T-2)\times m$ matrix of observations on the instruments.

Comparing

$$\widehat{\alpha}_{GMM} = (\Delta y_{-1}' Z W_N Z' \Delta y_{-1})^{-1} \Delta y_{-1}' Z W_N Z' \Delta y$$

and

$$\widehat{\alpha}_{AH} = (\Delta y_{-1}' Z(Z'Z)^{-1} Z' \Delta y_{-1})^{-1} \Delta y_{-1}' Z(Z'Z)^{-1} Z' \Delta y$$

note that for $T > 3$, there are two sources of greater (asymptotic) efficiency: $\widehat{\alpha}_{GMM}$ exploits more moment conditions ($m > 1$), and the 2SLS weight matrix ($W_N = (Z'Z)^{-1}$) is not optimal for the first-differenced specification under the maintained assumptions.

General results for GMM estimators indicate that $\widehat{\alpha}_{GMM}$ is strongly consistent (as $N \to \infty$ for fixed T) and asymptotically normal. For an arbitrary W_N

$$avar(\widehat{\alpha}_{GMM}) = N(\Delta y_{-1}' Z W_N Z' \Delta y_{-1})^{-1} \Delta y_{-1}' Z W_N \widehat{V}_N W_N Z' \Delta y_{-1} (\Delta y_{-1}' Z W_N Z' \Delta y_{-1})^{-1}$$

where

$$\widehat{V}_N = \frac{1}{N}\sum_{i=1}^{N}\left(Z_i' \widehat{\Delta v}_i \widehat{\Delta v}_i' Z_i\right)$$

and $\widehat{\Delta v}_{it} = \Delta y_{it} - \widehat{\alpha}\ \Delta y_{i,t-1}$ are first-differenced residuals, based on some consistent initial estimator $\widehat{\alpha}$.

The optimal (two step) GMM estimator thus sets $W_N = \widehat{V}_N^{-1}$, or

$$W_N = \left(\frac{1}{N}\sum_{i=1}^{N} Z_i' \widehat{\Delta v}_i \widehat{\Delta v}_i' Z_i\right)^{-1}$$

giving

$$avar(\widehat{\alpha}_{GMM}) = N(\Delta y_{-1}' Z W_N Z' \Delta y_{-1})^{-1}$$

For the special case in which $v_{it} \sim iid(0, \sigma_v^2)$, we can obtain a one step GMM estimator that is asymptotically equivalent to two step GMM.[3] For the first-differenced equations, this choice is not 2SLS, due to the serial correlation in Δv_{it} introduced by the first-differencing transformation.

While asymptotic results for the two step estimator only require an initial estimator that is consistent, small sample properties tend to be better when the estimate of the optimal weight matrix W_N stated above uses residuals $\widehat{\Delta v}_i$ based on an initial estimator that is also as efficient as possible.

Making explicit the dependence of the estimated optimal weight matrix on the initial consistent estimator

$$W_N(\widehat{\alpha}) = \left(\frac{1}{N}\sum_{i=1}^{N}\left(Z_i' \widehat{\Delta v}_i(\widehat{\alpha}) \widehat{\Delta v}_i(\widehat{\alpha})' Z_i\right)\right)^{-1}$$

indicates a small sample problem with the usual estimate of the asymptotic variance for the two step GMM estimator stated above. This neglects variation introduced by using an estimate $\widehat{\alpha}$ to construct the optimal weight matrix. In very large samples, this variation is negligible, and the usual expression for the asymptotic variance is correct. But in (reasonably large) finite samples, this additional variation makes two step inference based on $avar(\widehat{\alpha}_{GMM})$ unreliable. In fact, $avar(\widehat{\alpha}_{GMM})$ provides a good estimate of the variance of an infeasible GMM estimator, which uses the true value α rather than the initial estimate $\widehat{\alpha}$ to construct the optimal weight matrix.

Windmeijer (2005) proposes a finite sample correction that provides more accurate estimates of the variance of (linear) two step GMM estimators. t-tests based on these corrected standard errors are found to be as reliable as those based on the one step GMM estimator (where no parameters are estimated in the construction of the weight matrix).

[3] For each i, the $(T-2)\times(T-2)$ matrix $\widehat{\Delta v}_i \widehat{\Delta v}_i'$ in the expression for W_N is replaced by a matrix with main diagonal elements equal to 2, first off-diagonal elements equal to -1, and all other elements equal to zero. The variance of Δv_i is proportional to that known matrix in this special case.

The simple linear (generalised instrumental variables) expressions for these GMM estimators follows from the linearity of the moment conditions in α

$$\begin{aligned} E(y_{i,t-s}\Delta v_{it}) &= E(y_{i,t-s}(\Delta y_{it} - \alpha \Delta y_{i,t-1})) \\ &= E(y_{i,t-s}\Delta y_{it}) - \alpha E(y_{i,t-s}\Delta y_{i,t-1}) = 0 \end{aligned}$$

The 'standard assumptions' stated earlier imply a further $T - 3$ moment conditions that are quadratic in α. These can be written either as

$$E(\Delta v_{i,t-1}u_{it}) = 0 \ \text{ for } t = 4, \ldots, T$$

or as

$$E(\Delta u_{is}u_{iT}) = 0 \ \text{ for } s = 3, 4, \ldots, (T-1)$$

where $u_{it} = \eta_i + v_{it}$ is the error term for the untransformed equations in levels. Exploiting these additional moment conditions requires numerical optimisation procedures, and has been less common in practice. The resulting optimal non-linear GMM estimator is efficient for this model under the 'standard assumptions' stated previously, whereas the linear 'Arellano–Bond' GMM estimator is not. See Ahn and Schmidt (1995) for further discussion.

2.4. Weak instruments

Instrumental variables (and GMM) estimators have poor small sample properties in cases where the instruments, although valid, are only weakly correlated with the endogenous explanatory variables. This is relevant for the first-differenced GMM estimator in the AR(1) model in the case where $\alpha \to 1$. By analogy with random walks (innovations uncorrelated with past levels), the correlation between $\Delta y_{i,t-1}$ and the lagged levels $y_{i,t-s}$ for $s \geq 2$ becomes weaker as $\alpha \to 1$.

In the model

$$y_{it} = \alpha y_{i,t-1} + (\eta_i + v_{it})$$

α remains formally identified as $\alpha \to 1$, and the first-differenced GMM estimator remains consistent as $N \to \infty$, provided $E(\eta_i^2) \neq 0$. At $\alpha = 1$ we have

$$\Delta y_{i,t-1} = \eta_i + v_{i,t-1} \qquad \text{and} \qquad \Delta y_{i,t-2} = \eta_i + v_{i,t-2}$$

so that

$$E(\Delta y_{i,t-1}\Delta y_{i,t-2}) = E(\eta_i^2) \neq 0$$

$\Delta y_{i,t-2}$ is not a completely uninformative instrument for $\Delta y_{i,t-1}$ in the first-differenced equations.

Nevertheless, the Monte Carlo evidence presented in the Blundell and Bond (1998) paper, suggests that first-differenced GMM estimators become very imprecise, and subject to serious finite sample biases, for values of α around 0.8 and above, unless the available samples are huge. The finite sample bias is found to be downward, in the direction of the Within estimator, consistent with findings for 2SLS estimators in some simple cases where the weak instruments problem has been studied analytically.

Note also that if we combine

$$\begin{aligned} y_{it} &= \eta_i + \varepsilon_{it} \\ \varepsilon_{it} &= \alpha \varepsilon_{i,t-1} + v_{it} \end{aligned}$$

we obtain the alternative specification

$$y_{it} - \alpha y_{i,t-1} = \eta_i - \alpha \eta_i + \varepsilon_{it} - \alpha \varepsilon_{i,t-1}$$

or

$$y_{it} = \alpha y_{i,t-1} + (1-\alpha)\eta_i + v_{it}$$

In this specification, the process for y_{it} approaches a pure random walk as $\alpha \to 1$ (rather than a random walk with individual-specific drifts). Now at $\alpha = 1$, we have

$$\Delta y_{i,t-1} = v_{i,t-1}$$

Consequently lagged levels are completely uninformative instruments for $\Delta y_{i,t-1}$ in the limit case with $\alpha = 1$, and α is not identified using only the moment conditions

$$E(y_{i,t-s}\Delta v_{it}) = 0 \ \text{ for } t = 3, \ldots, T \text{ and } s \geq 2$$

for equations in first-differences. Although in this specification the OLS levels estimator is consistent when $\alpha = 1$. In this case a consistent test of the null hypothesis that $\alpha = 1$ can thus be obtained using a simple t-test based on the pooled OLS estimator of α.

3. Blundell-Bond estimators

3.1. 'System GMM'

Now consider the simple linear dynamic panel data model:

$$y_{it} = \alpha y_{i,t-1} + \eta_i + v_{it} \qquad |\alpha| < 1$$

for $i = 1, \ldots, N$ and $t = 2, \ldots, T$ under the assumptions stated above, and the additional (initial conditions) assumption

$$E(\Delta y_{i2}\eta_i) = 0.$$

This additional assumption has two implications.

(i) Now we have $E(\Delta y_{is}\eta_i) = 0$ for $s = 2, \ldots, T$, since the AR(1) specification implies

$$\begin{aligned}
\Delta y_{it} &= \alpha \Delta y_{i,t-1} + \Delta v_{it} = \alpha[\alpha \Delta y_{i,t-2} + \Delta v_{i,t-1}] + \Delta v_{it} \\
&= \alpha^2 \Delta y_{i,t-2} + \Delta v_{it} + \alpha \Delta v_{i,t-1} \\
&\ \ \vdots \\
&= \alpha^{t-2} \Delta y_{i2} + \sum_{s=0}^{t-3} \alpha^s \Delta v_{i,t-s} \quad \text{for } t = 3, \ldots, T
\end{aligned}$$

This then implies an additional $T-2$ non-redundant linear moment conditions for the equations in levels, which can be written (for example) as

$$E(\Delta y_{is} u_{iT}) = 0 \quad \text{for } s = 2, 3, \ldots, (T-1)$$

(ii) Given these additional linear moment restrictions, the quadratic moment restrictions are now redundant. For example, the product

$$\begin{aligned}
\Delta u_{is} u_{iT} &= (\Delta y_{is} - \alpha \Delta y_{i,s-1}) u_{iT} \\
&= \Delta y_{is} u_{iT} - \alpha \Delta y_{i,s-1} u_{iT}
\end{aligned}$$

so that $E(\Delta y_{is} u_{iT}) = 0$ and $E(\Delta y_{i,s-1} u_{iT}) = 0$ jointly imply $E(\Delta u_{is} u_{iT}) = 0$. Conveniently, the complete set of moment conditions implied by our standard assumptions and the initial conditions restriction $E(\Delta y_{i2}\eta_i) = 0$ can then be written as

$$E(y_{i,t-s} \Delta v_{it}) = 0 \quad \text{for } t = 3, \ldots, T \text{ and } s \geq 2$$

and

$$E(\Delta y_{is} u_{iT}) = 0 \quad \text{for } s = 2, 3, \ldots, (T-1)$$

and can thus be exploited using a linear GMM estimator.

However this is not just a matter of convenience. When this additional initial conditions assumption is valid, exploiting these additional moment conditions for the equations in levels can provide a dramatic improvement in efficiency, and reduction in finite sample bias, compared to the basic first-differenced GMM estimator, in cases where $\alpha \to 1$, or as the y_{it} series becomes more persistent. In this case the correlation between $\Delta y_{i,t-1}$ and lagged levels $y_{i,t-s}$ for $s \geq 2$ becomes weaker, and first-differenced GMM has poor finite sample properties associated with weak instruments — imprecise parameter estimates, and serious finite sample bias. In this context, exploiting the quadratic moment conditions could make a substantial improvement (Ahn and Schmidt, 1995). The Monte Carlo evidence presented in Blundell and Bond (1998) indicates that exploiting the additional linear moment conditions implied by this restriction on the initial conditions provides much more dramatic gains, provided that the additional initial conditions restriction is valid.

3.2. A restriction on the initial conditions?

The AR(1) specification determines y_{i2} given y_{i1}, so to guarantee that Δy_{i2} is uncorrelated with η_i we require a restriction on the behaviour of y_{i1}.

This is a form of stationarity restriction on the y_{it} series. The representation

$$\Delta y_{it} = \alpha^{t-2} \Delta y_{i2} + \sum_{s=0}^{t-3} \alpha^s \Delta v_{i,t-s} \quad \text{for } t = 3, \ldots, T$$

suggests (using backward recursion for earlier periods) that if the same model has generated the y_{it} series for long enough prior to our sample period, the observations on Δy_{it} would indeed be uncorrelated with η_i. 'Long enough' means long enough for any influence of the true start-up of the process to have become negligibly small (which in turn depends on the true value of α).

More formally, we can write

$$\begin{aligned} y_{i2} &= \alpha y_{i1} + \eta_i + v_{i2} \\ (y_{i2} - y_{i1}) &= (\alpha - 1)y_{i1} + \eta_i + v_{i2} \end{aligned}$$

and decompose

$$y_{i1} = \left(\frac{\eta_i}{1-\alpha}\right) + e_{i1}.$$

Then

$$\begin{aligned} \Delta y_{i2} &= (\alpha - 1)\left(\frac{\eta_i}{1-\alpha}\right) + (\alpha - 1)e_{i1} + \eta_i + v_{i2} \\ &= (\alpha - 1)e_{i1} + v_{i2}. \end{aligned}$$

The standard error components assumption implies $E(v_{i2}\eta_i) = 0$. A sufficient condition for $E(\Delta y_{i2}\eta_i) = 0$ is thus the restriction

$$E(e_{i1}\eta_i) = 0$$

3.3. Interpretation

Note that $\left(\frac{\eta_i}{1-\alpha}\right)$ is the level that our model specifies the y_{it} series will converge towards for individual i, if the process continues for long enough. $e_{i1} = y_{i1} - \left(\frac{\eta_i}{1-\alpha}\right)$ is the deviation from this convergent level at the start of our sample period. We require that these initial deviations are uncorrelated with η_i, or equivalently are uncorrelated with the convergent level itself. The initial observations y_{i1} can deviate randomly, but not systematically, from these convergent levels. This imposes a stationarity restriction on the mean of the y_{it} series, sometimes known as 'mean stationarity', but does not impose any restriction on the variance.

Whether this initial condition restriction is mild or strong will depend on the context, and particularly on the nature of the initial observations in our sample. As noted earlier, the restriction will hold automatically if the same process has generated the y_{it} series for long enough before the start of our sample period. Thus if we believe the AR(1) specification, and there is nothing special about our first observation period, it is reasonable to expect this restriction to hold. But if our first observation corresponds to the true start-up of the process, it may be an unreasonable restriction.

Computation of the extended (or 'system') GMM estimator is similar to the case discussed in Arellano and Bover (1995) in which suitably lagged first-differences of additional explanatory variables (x_{it}) can be used to obtain instruments for the equation(s) in levels. We add one (or more) equation(s) in levels to the set of first-differenced equations, for example

$$y_i^+ = \alpha y_{i(-1)}^+ + u_i^+$$

$$\begin{pmatrix} \Delta y_{i3} \\ \vdots \\ \Delta y_{iT} \\ y_{iT} \end{pmatrix} = \alpha \begin{pmatrix} \Delta y_{i2} \\ \vdots \\ \Delta y_{i,T-1} \\ y_{i,T-1} \end{pmatrix} + \begin{pmatrix} \Delta v_{i3} \\ \vdots \\ \Delta v_{iT} \\ \eta_i + v_{iT} \end{pmatrix}$$

and write the complete set of moment conditions as $E(Z_i^{+\prime} u_i^+) = 0$, where

$$Z_i^+ = \begin{pmatrix} Z_i & 0 & \dots & 0 \\ 0 & \Delta y_{i2} & \dots & \Delta y_{i,T-1} \end{pmatrix}$$

and Z_i was defined above.

Now defining $b_N(\alpha) = \frac{1}{N}\sum_{i=1}^{N} Z_i^{+\prime} u_i^+(\alpha)$ and choosing α to minimise $J_N(\alpha) = b_N(\alpha)' W_N b_N(\alpha)$ gives

$$\widehat{\alpha}_{GMM} = (y_{-1}^{+\prime} Z^+ W_N Z^{+\prime} y_{-1}^+)^{-1} y_{-1}^{+\prime} Z^+ W_N Z^{+\prime} y^+$$

for some positive definite weight matrix W_N.[4]

In general there is more than one equivalent representation of the available non-redundant moment conditions. We could use more moment conditions for the equations in levels, and fewer moment conditions for the equations in first-differences. However, we cannot express all the available moment conditions using only the equations in first-differences,

[4] When we combine equations in first-differences and equation(s) in levels, there is no special case with unobserved heterogeneity in the form of individual-specific *fixed* effects ($E(\eta_i^2) > 0$) in which $E((u_i^+)(u_i^+)^T)$ is proportional to a known matrix. As a result, there is no compelling choice of the weight matrix to obtain one step or initial consistent GMM estimators of this type. In practice it is advisable to assess the sensitivity of results to alternative reasonable choices.

or only the equations in levels. This is why we form an extended 'system' with both types of equations. Using all the available moment conditions for the equations in first-differences facilitates testing the validity of the additional moment conditions for the equations in levels (implied by the additional initial conditions restriction).

There are many extensions to the simple framework we have described here. The system GMM estimator extends to autoregressive models with low order MA and/or AR forms of serial correlation in the time-varying error component (v_{it}), and to models with additional explanatory variables (x_{it}), provided that the first-differences of these explanatory variables are uncorrelated with η_i. The Blundell and Bond (2000) paper on the estimation of production functions discusses an example with both these features, while Blundell et al. (2001) discusses conditions under which first-differences of the dependent variable are uncorrelated with η_i in linear dynamic models with additional explanatory variables. Note also that there is no requirement for a lagged dependent variable to be included in the model for these estimation methods to be useful.

The basic specification test for GMM estimators is the Sargan (1958) and Hansen (1982) test of overidentifying restrictions (or J test). An extension of this test procedure can be used to test nested hypotheses, where (only) a subset of the moment conditions that are valid under the null hypothesis remain valid under a less restrictive alternative hypothesis. Similarly, the Hausman test has the same sequential reasoning as the Difference Sargan/Hansen test, but focuses on the difference in the estimated parameter vectors under the null and under the alternative, rather than on the difference in the corresponding Sargan/Hansen test statistics. Arellano and Bond (1991) proposed a direct test for serial correlation in the residuals of the first-differenced equations. The Difference Sargan/Hansen and Hausman test procedures can also be used to test assumptions about the status of x_{it} explanatory variables, and to test the initial conditions restriction required for lagged values of Δy_{is} to be valid instruments in the levels equations.

4. Software

One important factor which has underpinned the popularity of Blundell–Bond estimators in applied work has been the availability of software which implements the method. Already in 1998, estimators which combined linear moment conditions for equations in levels with those for equations in first-differences were implemented in the open source DPD code for Gauss, written by Arellano and Bond and distributed through the IFS.[5] This was the code used to produce the empirical applications described in the Blundell and Bond (1998) paper, and in the subsequent Blundell and Bond (2000) paper on the estimation of production functions.

Access was extended beyond the community of Gauss users with the introduction of a DPD package for Oxmetrics, written by Jurgen Doornik and documented in Doornik et al. (1999). A landmark development was the introduction in 2003 of David Roodman's xtabond2 add-on command for Stata (see (Roodman, 2009)). This command allowed these extended or 'system' GMM estimators, incorporating moment conditions for equations in levels with suitably lagged first-differences of variables as the instruments, to be computed within the popular Stata environment. Another important feature of xtabond2 was the inclusion of the Windmeijer (2005) finite sample correction for the variance of two step linear GMM estimators, which enabled reliable inference to be based on the asymptotically efficient version of these estimators.[6]

The popularity of Roodman's xtabond2 command for Stata effectively put an end to development of the DPD code for Gauss. This command was soon incorporated in our teaching and used in our own empirical work, and remained the leading implementation of these estimators for many years. More recently, the introduction of Sebastian Kripfganz's xtdpdgmm add-on command for Stata has added an implementation of estimators which incorporate the non-linear moment conditions discussed in Ahn and Schmidt (1995), as an alternative to the linear moment conditions for equations in levels discussed in Arellano and Bover (1995) and Blundell and Bond (1998).[7] At the time of writing, the xtdpdgmm command for Stata would be our recommended implementation of these extended GMM estimators for dynamic panel data models.[8]

5 The earliest version was documented in Arellano and Bond (1988). The latest version was documented in Arellano and Bond (1998), and is still available from the IFS web site.

6 Neither 'system' GMM estimators nor Windmeijer-corrected standard errors were available in the first version of Stata's official xtabond command, although these features have since been added to xtabond, and to Stata's later xtdpd and xtdpdsys commands.

7 See Kripfganz (2019). The xtdpdgmm command also corrects some known bugs in the xtabond2 command, which can affect the reported degrees of freedom for tests of overidentifying restrictions in some circumstances. Other recent developments include the introduction of commands which implement these estimators in R, including the pgmm function within the plm package (Croissant and Millo, 2019), and the pdynmc package (Fritsch et al., 2021).

8 We remain less keen on the default choice for the one step weight matrix used by both xtdpdgmm and xtabond2 when equations in levels are combined with equations in first-differences. This weight matrix is asymptotically efficient only in the absence of a time-invariant unobserved component (η_i) in the error term of the equations in levels, while it is the presence of such unobserved heterogeneity which motivates the development and use of these GMM estimators for panel data. However we appreciate that both these commands allow the sensitivity of the results to alternative choices for the one step weight matrix to be assessed.

5. Summary and conclusions

In this commentary we have placed the Blundell and Bond (1998) paper in the context of the early development of panel data estimators that accounted for dynamics and persistent economic series. The ideas grew out of the increasing empirical interest in panel data models of individual and firm behaviour. The initial work focused on appropriate econometric methods to estimate dynamic models using unbalanced panel data with many firms and/or individuals but covering a small number of time periods. A natural approach noted that unobserved firm-specific 'fixed' effects could be eliminated by taking first-differences, while using suitably lagged values of the dependent variable, and of endogenous or predetermined explanatory variables, as instruments. Generalised Method of Moments (GMM) provided a natural framework to obtain asymptotically efficient estimators based on these moment conditions. This approach was popularised by Arellano and Bond (1991) and their DPD software. The breakthrough in Arellano and Bover (1995) was to show that, under certain conditions, suitably lagged first-differences of explanatory variables are available as instruments for the equations in levels. One contribution of our paper was to derive the conditions under which first-differences of the dependent variable would or would not be uncorrelated with individual-specific 'fixed' effects in autoregressive models. Another influential contribution was to highlight both the magnitude of the bias when the first-differenced GMM estimator is used to estimate autoregressive models for highly persistent series, and the potential to reduce that bias by using additional valid moment conditions for the equations in levels — thereby popularising the use of these extended or 'system' GMM estimators.

References

Ahn, S.C., Schmidt, P., 1995. Efficient estimation of models for dynamic panel data. J. Econometrics 68, 5–28.
Anderson, T.W., Hsiao, C., 1981. Estimation of dynamic models with error components? J. Amer. Statist. Assoc. 76, 598–606.
Arellano, M., Bond, S.R., 1988. Dynamic Panel Data Estimation using DPD - a Guide for Users. IFS Working Paper no. 88/15.
Arellano, M., Bond, S.R., 1991. Some tests of specification for panel data: Monte Carlo evidence and an application to employment equations. Rev. Econom. Stud. 58, 277–297.
Arellano, M., Bond, S.R., 1998. Dynamic panel data estimation using DPD98 for GAUSS - a guide for users. https://ifs.org.uk/publications/dpd-gauss.
Arellano, M., Bover, O., 1995. Another look at the instrumental-variable estimation of error-components models. J. Econometrics 68, 29–52.
Blundell, R.W., Bond, S.R., 1998. Initial conditions and moment restrictions in dynamic panel data models. J. Econometrics 87, 115–143.
Blundell, R.W., Bond, S.R., 2000. GMM estimation with persistent panel data: an application to production functions. Econometric Rev. 19, 321–340.
Blundell, R.W., Bond, S.R., Devereux, M.P., Schiantarelli, F., 1992. Investment and Tobin's Q: evidence from company panel data. J. Econometrics 51, 233–257.
Blundell, R.W., Bond, S.R., Windmeijer, F., 2001. Estimation in dynamic panel data models: improving on the performance of the standard GMM estimator. In: Baltagi, B.H., Fomby, T.B., Carter-Hill, R. (Eds.), Nonstationary Panels, Panel Cointegration, and Dynamic Panels (Advances in Econometrics, 15). Emerald Group Publishing Limited, Bingley, pp. 53–91.
Blundell, R.W., Smith, R.J., 1991. Conditions initiales et estimation efficace dans les modeles dynamiques sur donnees de panel:une application au comportement d'investissement des entreprises. Anna. D'Econom. Stat. (20/21), 109–123.
Croissant, Y., Millo, G., 2019. Panel Data Econometrics with R. John Wiley & Sons, Hoboken NJ.
Doornik, J.A., Arellano, M., Bond, S.R., 1999. Panel data estimation using DPD for Ox. https://www.doornik.com/download/oxmetrics7/Ox_Packages/dpd.pdf.
Fritsch, M., Pua, A.A.Y., Schnurbus, J., 2021. pdynmc: a package for estimating linear dynamic panel data models based on nonlinear moment conditions. R J. 13, 218–231.
Hansen, L.P., 1982. Large sample properties of generalised method of moment estimators. Econometrica 50, 1029–1054.
Hayashi, F., 1982. Tobin's marginal q and average q: a neoclassical interpretation. Econometrica 50, 213–224.
Holtz-Eakin, D., Newey, W., Rosen, H.S., 1988. Estimating vector autoregressions with panel data. Econometrica 56 (6), 1371–1395.
Kripfganz, S., 2019. Generalized method of moments estimation of linear dynamic panel data models. In: Proceedings of the 2019 London Stata Conference.
Nickell, S., 1981. Biases in dynamic models with fixed effects. Econometrica 49 (6), 1417–1426.
Roodman, D.M., 2009. How to do xtabond2: an introduction to difference and system GMM in stata. Stata J. 9, 86–136.
Sargan, J.D., 1958. The estimation of economic relationships using instrumental variables. Econometrica 26 (3), 393–415.
Windmeijer, F., 2005. A finite sample correction for the variance of linear efficient two-step GMM estimators. J. Econometrics 126, 25–51.