Port Econ J (2002) 1: 141–162

Portuguese Economic Journal

Dynamic panel data models: a guide to micro data methods and practice*

Stephen R. Bond

Nuffield College, Oxford, UK, and Institute for Fiscal Studies, 7 Ridgmount Street, London, WC1E 7AE, UK (e-mail: steve.bond@nuffield.ox.ac.uk)

Abstract. This paper reviews econometric methods for dynamic panel data models, and presents examples that illustrate the use of these procedures. The focus is on panels where a large number of individuals or firms are observed for a small number of time periods, typical of applications with microeconomic data. The emphasis is on single equation models with autoregressive dynamics and explanatory variables that are not strictly exogenous, and hence on the Generalised Method of Moments estimators that are widely used in this context. Two examples using firm-level panels are discussed in detail: a simple autoregressive model for investment rates; and a basic production function.

Key words: Panel data – Dynamic models – Generalised method of moments

JEL Classification: C23

1 Introduction

Panel data is now widely used to estimate dynamic econometric models. Its advantage over cross-section data in this context is obvious: we cannot estimate dynamic models from observations at a single point in time, and it is rare for single cross-section surveys to provide sufficient information about earlier time periods for dynamic relationships to be investigated. Its advantages over aggregate time series data include the possibility that underlying microeconomic dynamics may be

* This paper is prepared for a special issue of the Portuguese Economic Journal. My understanding of this subject has benefitted immeasurably from the input of colleagues, notably Manuel Arellano, Richard Blundell, Costas Meghir, Steve Nickell and Frank Windmeijer. I thank Joao Santos Silva, Frank Windmeijer and other participants at the Cemmap/ESRC Econometric Study Group workshop in London on 22 February 2002 for detailed comments on an earlier draft. Financial support from the ESRC Centre for the Microeconomic Analysis of Public Policy is gratefully acknowledged.

obscured by aggregation biases,[1] and the scope that panel data offers to investigate heterogeneity in adjustment dynamics between different types of individuals, household or firms. Whilst these advantages are shared by repeated cross-section or cohort data, from which pseudo-panel data on grouped observations can be constructed,[2] genuine panel data - with repeated observations on the *same* individuals - will typically allow more of the variation in the micro data to be used in constructing parameter estimates, as well as permitting the use of relatively simple econometric techniques.

Dynamic models are of interest in a wide range of economic applications, including Euler equations for household consumption, adjustment cost models for firms' factor demands, and empirical models of economic growth. Even when coefficients on lagged dependent variables are not of direct interest, allowing for dynamics in the underlying process may be crucial for recovering consistent estimates of other parameters. An example occurs in the estimation of production functions when productivity shocks are serially correlated and relative factor inputs respond to these shocks; this is discussed further in Section 3 below.

The focus in this paper will be on the estimation of single equation, autoregressive-distributed lag models from panels with a large number of cross-section units, each observed for a small number of time periods. This situation is typical of micro panel data on individuals or firms, and calls for estimation methods that do not require the time dimension to become large in order to obtain consistent parameter estimates. Assumptions about the properties of initial conditions also play an important role in this setting, since the influence of the initial observations on each subsequent observation cannot safely be ignored when the time dimension is short. We also focus on methods that can be used in the absence of any strictly exogenous explanatory variables or instruments, and that extend easily to models with predetermined or endogenous explanatory variables. Strict exogeneity rules out any feedback from current or past shocks to current values of the variable, which is often not a natural restriction in the context of economic models relating several jointly determined outcomes, such as consumption and income or investment and Tobin's q. Identification then depends on limited serial correlation in the error term of the equation, which leads to a convenient and widely used class of Generalised Method of Moments (GMM) estimators for this type of dynamic panel data model.

Rigorous surveys of these estimators can be found in, for example, Arellano and Honore (2001) or Blundell, Bond and Windmeijer (2000). The emphasis here will be on an intuitive review of these methods, intended to give the applied researcher an appreciation for when it may be reasonable to use particular GMM estimators, and how this can be evaluated in practice. This will be illustrated in the context of two empirical examples. The first uses panel data on UK firms to estimate a simple autoregressive model for company investment rates. The second uses panel data on US firms to estimate a basic production function specification. An obvious difference between these two examples is that the former illustrates the use of these methods to estimate a univariate dynamic model, whilst the latter considers

[1] See, for example, Nickell (1986).

[2] See, for example,Verbeek (1996).

a model with additional explanatory variables. A more subtle difference is that the output and input series used in the production function context are near unit root series, whilst the investment rates - which are basically growth rates in the capital stock - are not. Whilst the time series properties of the series are not crucial for the asymptotic distribution theory in this setting, where the number of cross-section units is assumed to become large with the number of time periods treated as fixed, they can nevertheless be crucial for the identification of parameters of interest and for the finite sample properties of particular GMM estimators. Specifically, estimators which rely on first-differencing or related transformations to eliminate unobserved individual-specific effects, and then use lagged values of endogenous or predetermined variables as instruments for subsequent first-differences, can be expected to perform poorly in situations where the series are close to being random walks, so that their first-differences are close to being innovations, and will not identify parameters of interest in some limiting cases.

Section 2 discusses the estimation of simple autoregressive models and illustrates the use of these first-differenced GMM methods in a setting where this identification problem does not arise. Section 3 discusses the use of additional moment conditions that typically require stronger assumptions on the initial conditions, but which can be highly informative in cases where identification using the first-differenced equations alone becomes weak. This is illustrated in a multivariate context. Section 4 concludes.

2 Autoregressive models

In this section we focus on estimation methods for the simple $AR(1)$ model

$$y_{it} = \alpha y_{i,t-1} + (\eta_i + v_{it});\ |\alpha| < 1;\ i = 1, 2, \ldots, N;\ t = 2, 3, \ldots, T \qquad (2.1)$$

where y_{it} is an observation on some series for individual i in period t, $y_{i,t-1}$ is the observation on the same series for the same individual in the previous period, η_i is an unobserved individual-specific time-invariant effect which allows for heterogeneity in the means of the y_{it} series across individuals, and v_{it} is a disturbance term. A key assumption we maintain throughout is that the disturbances v_{it} are independent across individuals. The number of individuals for which data is available (N) is assumed to be large whilst the number of time periods for which data is available (T) is assumed to be small, and asymptotic properties are considered as N becomes large with T fixed.[3]

We treat the individual effects (η_i) as being stochastic, which here implies that they are necessarily correlated with the lagged dependent variable $y_{i,t-1}$ unless the distribution of the η_i is degenerate. Initially we further assume that the disturbances (v_{it}) are serially uncorrelated. These jointly imply that the Ordinary Least Squares (OLS) estimator of α in the levels equations (2.1) is inconsistent, since the

[3] In this context it is straightforward to allow for time-specific effects, common to all individuals, simply by including period-specific intercepts in the specification, or by transforming the series into deviations from period-specific means (i.e. the mean across all individuals for a particular time period). We simplify our discussion by not dealing explicitly with this case.

explanatory variable $y_{i,t-1}$ is positively correlated with the error term $(\eta_i + v_{it})$ due to the presence of the individual effects, and this correlation does not vanish as the number of individuals in the sample gets larger.[4] Standard results for omitted variable bias indicate that, at least in large samples, the OLS levels estimator is biased upwards.

The Within Groups estimator eliminates this source of inconsistency by transforming the equation to eliminate η_i. Specifically the mean values of y_{it}, $y_{i,t-1}$, η_i and v_{it} across the $T-1$ observations for each individual i are obtained, and the original observations are expressed as deviations from these individual means. OLS is then used to estimate these transformed equations. Since the mean of the time-invariant η_i is itself η_i, these individual effects are removed from the transformed equations. However, for panels where the number of time periods available is small, this transformation induces a non-negligible correlation between the transformed lagged dependent variable and the transformed error term. The transformed lagged dependent variable is $y_{i,t-1} - \frac{1}{T-1}(y_{i1} + \ldots + y_{it} + \ldots + y_{i,T-1})$ whilst the transformed error term is $v_{it} - \frac{1}{T-1}(v_{i2} + \ldots + v_{i,t-1} + \ldots + v_{iT})$. The component $\frac{-y_{it}}{T-1}$ in the former is correlated with v_{it} in the latter, and the component $\frac{-v_{i,t-1}}{T-1}$ in the latter is correlated with $y_{i,t-1}$ in the former. These leading correlations, which are both negative, dominate positive correlations between other components such as $\frac{-v_{i,t-1}}{T-1}$ and $\frac{-y_{i,t-1}}{T-1}$, so that the correlation between the transformed lagged dependent variable and the transformed error term can be shown to be negative.[5] This correlation does not vanish as the number of individuals in the sample increases, so that the Within Groups estimator is also inconsistent here.[6] Standard results for omitted variables bias indicate that, at least in large samples, the Within Groups estimator is biased downwards.[7]

The fact that these two estimators are likely to be biased in opposite directions is useful. Thus we might hope that a candidate consistent estimator will lie between the OLS and Within Groups estimates, or at least not be significantly higher than the former or significantly lower than the latter. In cases where the $AR(1)$ model seems well specified and this pattern is not observed, we might suspect either inconsistency or severe finite sample bias for the supposedly consistent estimator, and go on to consider more rigorous testing to investigate if this is indeed the case.[8]

---

[4] Nor does this correlation vanish as the number of time periods increases, so that OLS levels remains inconsistent for panels with large T.

[5] See Nickell (1981).

[6] However the contribution of each time period to the individual means becomes negligibly small as the number of time periods gets larger. Consequently this correlation induced by the transformation vanishes, and the Within Groups estimator is consistent in the case of large T panels.

[7] The lagged dependent variable in the $AR(1)$ model (2.1) is an example of an explanatory variable that is predetermined with respect to the disturbances (v_{it}), but not strictly exogenous in the sense of being uncorrelated with all past, present and future values of these disturbances. The Within Groups estimator, which introduces all realisations of the v_{it} series into the transformed error term, is only consistent in large N, fixed T panels if all the right-hand side variables are strictly exogenous.

[8] In cases where the $AR(1)$ model does not seem well specified, a similar indication may be available by considering the sum of the estimated coefficients on the lagged values of the dependent variable in higher-order autoregressive models. For example, in the $AR(2)$ model $y_{it} = \alpha_1 y_{i,t-1} + \alpha_2 y_{i,t-2} + (\eta_i + v_{it})$, we can compare OLS levels and Within Groups estimates of $(\alpha_1 + \alpha_2)$.

Maximum Likelihood estimators for the $AR(1)$ panel data model have been developed, but an awkward feature of dynamic models in short T panels is that the distribution of y_{it} for $t = 2, 3, \ldots, T$ depends in a non-negligible way on what is assumed about the distribution of the initial conditions y_{i1}. In specifying the model (2.1) we have not restricted the process that generates these initial conditions, and a wide variety of alternative processes could be considered. For example the y_{i1} may be stochastic or non-stochastic, correlated or uncorrelated with the individual effects η_i, specified such that the mean of the y_{it} series for individual i takes a stationary value $(\frac{\eta_i}{1-\alpha})$ for each observation $t = 1, 2, \ldots, T$, or to satisfy higher order stationarity properties. Different assumptions about the nature of the initial conditions will lead to different likelihood functions, and the resulting Maximum Likelihood estimators of α can be inconsistent when this initial conditions process is mis-specified. Interested readers are referred to Hsiao (1986) for further discussion.[9]

Instrumental Variables estimators which require much weaker assumptions about the initial conditions have therefore been attractive in this context. The basic first-differenced Two Stage Least Squares (2SLS) estimator for the $AR(1)$ panel data model was proposed by Anderson and Hsiao (1981, 1982), initially as a way of obtaining a consistent starting value for computation of Maximum Likelihood estimators. The first-differencing transformation also eliminates the individual effects η_i from the model

$$\Delta y_{it} = \alpha \Delta y_{i,t-1} + \Delta v_{it};\ |\alpha| < 1;\ i = 1, 2, \ldots, N;\ t = 3, 4, \ldots, T \qquad (2.2)$$

where $\Delta y_{it} = y_{it} - y_{i,t-1}$. An important difference from the Within transformation, however, is that first-differencing does not introduce all realisations of the disturbances $(v_{i2}, v_{i3}, \ldots, v_{iT})$ into the error term of the transformed equation for period t. The dependence of Δv_{it} on $v_{i,t-1}$ implies that OLS estimates of α in the first-differenced model are inconsistent, with the direction of the inconsistency being downward and typically greater than that found for the Within Groups estimator.[10] However consistent estimates of α can now be obtained using 2SLS with instrumental variables that are both correlated with $\Delta y_{i,t-1}$ and orthogonal to Δv_{it}.

The only assumption required on the initial conditions y_{i1} is that they are uncorrelated with the subsequent disturbances v_{it} for $t = 2, 3, \ldots, T$, in which case the initial conditions are said to be predetermined. The correlation between y_{i1} and the individual effects η_i is left unrestricted, and there is no requirement for any stationarity condition to be satisfied. Together with the previous assumption that the disturbances v_{it} are serially uncorrelated, predetermined initial conditions imply that the lagged level $y_{i,t-2}$ will be uncorrelated with Δv_{it} and thus available as an

---

[9] Blundell and Smith (1991) develop a class of maximum likelihood estimators which avoid this fragility by conditioning on the initial observations in the sample. Binder, Hsiao and Pesaran (2000) present maximum likelihood estimators which allow the possibility of $\alpha = 1$ and hence can be considered for unit root testing, although specific assumptions about the initial conditions are still required in the $\alpha < 1$ case. Both these approaches can be extended to vector autoregressive models, although not, so far as I am aware, to models where the process for additional explanatory (x) variables is not specified.

[10] The transformations, and hence the corresponding OLS estimators, coincide in the special case when $T = 3$.

instrumental variable for the first-differenced equations in (2.2). The resulting 2SLS estimator is consistent in large N, fixed T panels, and identifies the autoregressive parameter α provided at least 3 time series observations are available ($T \geqslant 3$).[11]

Additional instruments are available when the panel has more than 3 time series observations. Whilst y_{i1} is the only instrument that can be used in the first-differenced equation for period $t = 3$, both y_{i1} and y_{i2} can be used in the first-differenced equation for period $t = 4$, and the vector $(y_{i1}, y_{i2}, \ldots, y_{i,T-2})$ can be used in the first-differenced equation for period $t = T$. Since the model is overidentified with $T > 3$, and the first-differenced error term Δv_{it} has a first-order moving average form of serial correlation if the maintained assumption that the v_{it} are serially uncorrelated is correct, 2SLS is not asymptotically efficient even if the complete set of available instruments is used for each equation and the disturbances v_{it} are homoskedastic. The Generalised Method of Moments (GMM), developed by Hansen (1982), provides a convenient framework for obtaining asymptotically efficient estimators in this context, and first-differenced GMM estimators for the $AR(1)$ panel data model were developed by Holtz-Eakin, Newey and Rosen (1988) and Arellano and Bond (1991). Essentially these use an instrument matrix of the form

$$Z_i = \begin{bmatrix} y_{i1} & 0 & 0 & \ldots & 0 & \ldots & 0 \\ 0 & y_{i1} & y_{i2} & \ldots & 0 & \ldots & 0 \\ . & . & . & \ldots & . & \ldots & . \\ 0 & 0 & 0 & \ldots & y_{i1} & \ldots & y_{i,T-2} \end{bmatrix}, \tag{2.3}$$

where rows correspond to the first-differenced equations for periods $t = 3, 4, \ldots, T$ for individual i, and exploit the moment conditions

$$E\left[Z_i' \Delta v_i\right] = 0 \quad \text{for } i = 1, 2, \ldots, N \tag{2.4}$$

where $\Delta v_i = (\Delta v_{i3}, \Delta v_{i4}, \ldots, \Delta v_{iT})'$.

In general, the asymptotically efficient GMM estimator based on this set of moment conditions minimises the criterion

$$J_N = \left(\frac{1}{N} \sum_{i=1}^{N} \Delta v_i' Z_i \right) W_N \left(\frac{1}{N} \sum_{i=1}^{N} Z_i' \Delta v_i \right) \tag{2.5}$$

using the weight matrix

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

where the $\widehat{\Delta v}_i$ are consistent estimates of the first-differenced residuals obtained from a preliminary consistent estimator. Hence this is known as a two-step GMM estimator. Under homoskedasticity of the v_{it} disturbances, the particular structure

[11] This 2SLS estimator is also consistent in large T panels, although as we have seen the Within Groups estimator is also consistent in this context.

of the first-differenced model implies that an asymptotically equivalent GMM estimator can be obtained in one-step, using instead the weight matrix

$$W_{1N} = \left[\frac{1}{N} \sum_{i=1}^{N} (Z_i' H Z_i) \right]^{-1},$$

where H is a $(T-2)$ square matrix with 2's on the main diagonal, -1's on the first off-diagonals and zeros elsewhere. Notice that W_{1N} does not depend on any estimated parameters.

At a minimum this suggests that the one-step estimator using W_{1N} is a reasonable choice for the initial consistent estimator used to compute the optimal weight matrix W_N and hence to compute the two-step estimator. In fact a lot of applied work using these GMM estimators has focused on results for the one-step estimator rather than the two-step estimator. This is partly because simulation studies have suggested very modest efficiency gains from using the two-step version, even in the presence of considerable heteroskedasticity,[12] but more importantly because the dependence of the two-step weight matrix on estimated parameters makes the usual asymptotic distribution approximations less reliable for the two-step estimator. Simulation studies have shown that the asymptotic standard errors tend to be much too small, or the asymptotic t-ratios much too big, for the two-step estimator, in sample sizes where the equivalent tests based on the one-step estimator are quite accurate.[13] Windmeijer (2000) provides a formal analysis of this issue, and proposes a finite-sample correction for the asymptotic variance of the two-step GMM estimator which is potentially very useful in this class of models.

When $T > 3$ and the model is overidentified, the validity of the assumptions used to obtain (2.4) can be tested using the standard GMM test of overidentifying restrictions, or Sargan test. In particular NJ_N in (2.5) has an asymptotic χ^2 distribution under the null that these moment conditions are valid (see Sargan, 1958, 1988; Hansen, 1982). In this context the key identifying assumption that there is no serial correlation in the v_{it} disturbances can also be tested by testing for no second-order serial correlation in the first-differenced residuals; see Arellano and Bond (1991).[14] Bond and Windmeijer (2002) review alternative approaches to testing hypotheses about the parameter of interest α.

These methods extend straightforwardly to higher order autoregressive models and to models with limited moving-average serial correlation in the disturbances v_{it}, provided that a minimum number of time series observations are available to identify the parameters of interest. For example, if v_{it} is (2.1) is itself $MA(1)$, then the first-differenced error term Δv_{it} in (2.2) is $MA(2)$. As a result $y_{i,t-2}$ would not be a valid instrumental variable in these first-differenced equations, but $y_{i,t-3}$ and longer lags remain available as instruments. In this case α is identified using this restricted

---

[12] See, for example, Arellano and Bond (1991), Blundell and Bond (1998) and Blundell, Bond and Windmeijer (2000).

[13] See also Bond and Windmeijer (2002).

[14] Negative first-order serial correlation is expected in the first-differenced residuals if the v_{it} are serially uncorrelated. Positive serial correlation is expected in the levels residuals due to the presence of the individual effects η_i.

instrument set provided $T \geqslant 4$. Extensions to models with additional explanatory (x) variables are also straightforward, and are discussed briefly in the next section. Moreover whilst we have introduced these GMM estimators for the case where first-differencing is used to eliminate the individual effects, the two key properties of the first-differencing transformation needed here - eliminating the time-invariant individual effects whilst not introducing disturbances for periods earlier than period $t-1$ into the transformed error term - are shared by a wide class of alternative transformations. Arellano and Bover (1995) show that numerically identical GMM estimators can be obtained using any of the alternative transformations in this class.[15]

The first-differenced GMM estimators described here are straightforward to compute, as the moment conditions in (2.4) are all linear in the parameter α.[16] Under the very mild assumption that the individual effects η_i and the disturbances v_{it} are uncorrelated, additional moment conditions are available for this model which are quadratic in α and hence require iterative computational procedures. Additional linear moment conditions are available under a homoskedasticity assumption. Ahn and Schmidt (1995) characterise the additional moment conditions in both these cases. Perhaps more useful in practice, however, are additional linear moment conditions which result from assuming that the initial conditions satisfy mean stationarity, so that the series $(y_{i1}, y_{i2}, \ldots, y_{iT})$ have a constant mean $\frac{\eta_i}{1-\alpha}$ for each individual i. These are discussed in the next section, after we illustrate the estimators discussed here for the $AR(1)$ model using data on investment rates for a panel of UK firms.

2.1 Example: company investment rates

Table 1 reports estimates of the simple $AR(1)$ specification for data on investment rates for the panel of UK companies used in Bond, Klemm, Newton-Smith, Syed and Vlieghe (2002). Since the investment data (I_{it}) measure gross investment expenditures and the capital stock series (K_{it}) is a measure of the net capital stock at replacement cost, the investment rates $y_{it} = (I_{it}/K_{it})$ give an approximate measure of the growth rate in the net capital stock plus a depreciation rate. The model estimated is a first-order autoregressive specification as in equation (2.1), with year-specific intercepts (c_t) included to account for common cyclical or trend components in these investment rates

$$\left(\frac{I_{it}}{K_{it}}\right) = c_t + \alpha\left(\frac{I_{i,t-1}}{K_{i,t-1}}\right) + (\eta_i + v_{it}). \tag{2.6}$$

Variation across firms in depreciation rates provides one motivation for suspecting the presence of individual-specific effects in this context. The sample contains 703

[15] The corresponding estimators are numerically identical when the panel is balanced (all individuals have complete data on y_{it} for $t = 1, 2, \ldots, T$) and all the available linear moment conditions are exploited. Alternative transformations may lead to more or less efficient estimators when only a subset of the available instruments are used.

[16] Computer programmes that include these estimators include DPD98 for Gauss (Arellano and Bond, 1998), Ox (www.oxmetrics.net), PcGive (www.pcgive.com) and Stata (www.stata.com).

publicly traded UK firms for which we have consecutive annual data from published company accounts for a minimum of 4 years between 1987 and 2000. Further details of the sample and the data are provided in Bond et al. (2002).

The first two columns of Table 1 report OLS levels and Within Groups estimates of the parameter α, together with heteroskedasticity-consistent estimates of the asymptotic standard errors.[17] Recall that the OLS estimate is likely to be biased upwards and the Within Groups estimate is likely to be biased downwards if the $AR(1)$ model in (2.6) provides a good representation for this series. Certainly the OLS estimate is considerably higher than the Within Groups estimate. The serial correlation tests $m1$ and $m2$ reported for OLS levels test the null hypotheses of no first-order serial correlation and no second-order serial correlation respectively in the OLS residuals $y_{it} - \widehat{\alpha} y_{i,t-1}$, where $\widehat{\alpha}$ is the OLS estimate of α. The test statistics are asymptotically standard normal under the null of no serial correlation, and their signs indicate the sign of the estimated autocorrelation coefficients in the residuals.[18] Whilst we might expect to find significant positive serial correlation at both first order and second order in consistent estimates of these levels residuals, due to the presence of individual effects, there is no reason to expect this pattern if the OLS estimate of α, and hence these estimates of the residuals, are seriously biased. For the Within Groups estimator, the reported test statistics consider serial correlation in the first-differenced residuals, $\Delta y_{it} - \widehat{\alpha} \Delta y_{i,t-1}$, where $\widehat{\alpha}$ is here the Within Groups estimate of α. Similarly, however, the finding of significant second-order serial correlation in these estimates of the residuals need not indicate that the $AR(1)$ model is mis-specified, since this estimate of α and hence these estimates of the first-differenced residuals are likely to be biased.

The third column reports the simple, just-identified 2SLS estimator for the equations in first-differences, using $y_{i,t-2}$ as the instrumental variable. This gives an estimate of α which is significantly positive, but well below the OLS estimate and well above the Within Groups estimate. Whilst we cannot use the Sargan statistic to test the validity of overidentifying restrictions for this just-identified estimator, the pattern of serial correlation in the first-differenced residuals is consistent with the maintained assumption that the v_{it} disturbances are serially uncorrelated, so that Δv_{it} should have significant negative first-order serial correlation but no significant second-order serial correlation. We have some evidence that the $AR(1)$ model is well specified for this series, and the ranking of the OLS, 2SLS and Within Groups parameter estimates is in line with what we would expect if this is the case *and* the consistent 2SLS estimator is not subject to any serious finite sample bias in this context.

The fourth column reports a one-step first-differenced GMM estimator for this specification, using only a subset of the potentially available instrument matrix Z_i in (2.3). In particular, only instruments corresponding to $y_{i,t-2}$ and $y_{i,t-3}$ are used here, and columns of Z_i containing longer lags for the later first-differenced

[17] All computations are done using DPD98 for Gauss; see Arellano and Bond (1998).

[18] These test statistics are standardised first-order and second-order residual autocovariances; see Arellano and Bond (1991) for details.

equations are omitted.[19] We obtain a modest improvement in the precision of the parameter estimate from the addition of these additional moment conditions. As this model is now overidentified, we can use the Sargan statistic to test the validity of the overidentifying restrictions. In this case we obtain a $\chi^2(22)$ statistic of 23.84, giving the reported p-value of 0.36.[20] Consistent with the evidence from the serial correlation tests, the null hypothesis that these moment conditions are invalid is not rejected at any conventional significance level.

The final column reports the one-step first-differenced GMM estimator which uses the complete set of linear moment restrictions implied by our assumptions that the v_{it} disturbances are serially uncorrelated and the initial conditions y_{i1} are predetermined.[21] We obtain a further very modest improvement in the precision of the parameter estimate, and again there is no evidence either from the serial correlation tests or from the Sargan test that the simple $AR(1)$ model is mis-specified for this series.[22] Noteworthy is what happens in this case if we consider the two-step version of the GMM estimator. This gives an estimate of 0.1806, with an asymptotic standard error of 0.0227. Whilst the point estimate is thus similar to the one-step estimate of 0.1560 reported in the final column of Table 1, there is apparently a reduction of some 30% in the standard error of this estimate. Whilst some gain in precision could be expected in the presence of heteroskedastic disturbances, no improvement of this magnitude has ever been reported in simulation studies for these estimators. Using the finite sample correction proposed by Windmeijer (2000) gives a corrected standard error of 0.0314 for this two-step estimate, which is only marginally lower than the standard error for the one-step estimate reported in Table 1, and much more in line with the relative efficiency of these two estimators reported in simulations. Several studies have noted that inference based on the conventional asymptotic variance matrix for the two-step GMM estimator can be highly inaccurate in this context; see, for example, Bond and Windmeijer (2002).

---

[19] Notice however that this estimator still uses many more instrumental variables than the just-identified 2SLS estimator in column three. The instrument matrix for 2SLS here is the single column vector $Z_i = (y_{i1}, y_{i2}, \ldots, y_{i,T-2})'$, whilst even the restricted version of (2.3) used in column four has 23 columns.

[20] The Sargan test reported here is based on the minimised value of the associated two-step GMM estimator, which has a convenient asymptotic χ^2 distribution regardless of heteroskedasticity. Notice that these one-step and two-step GMM estimators are based on the same set of moment conditions, so that the overidentifying restrictions tested here are those used by both estimators.

[21] The instrument matrix Z_i has 78 columns in this case.

[22] One reason to compare the estimator based on the full set of instruments in column five with the estimator based on the restricted instrument set in column four is that GMM estimators based on too many moment conditions can be subject to potentially severe overfitting biases in small samples. This does not seem to be an issue here. The loss of relevant information caused by omitting the more distant lags as instruments will often be very modest, as we find here. Bowsher (2000) also notes that the power of the Sargan test to detect invalid overidentifying restrictions can decline dramatically in finite samples if an excessive number of moment conditions are used.

Table 1. Alternative estimates of the $AR(1)$ specification for company investment rates

Dependent variable: $(I/K)_t$					
	OLS levels	Within groups	2SLS DIF	GMM DIF	GMM DIF
$(I/K)_{t-1}$	0.2669	−0.0094	0.1626	0.1593	0.1560
	(.0185)	(.0181)	(.0362)	(.0327)	(.0318)
m1	−4.71	−11.36	−10.56	−10.91	−11.12
m2	2.52	−2.02	0.61	0.52	0.46
Sargan				.36	.43
Instruments			$(I/K)_{t-2}$	$(I/K)_{t-2}$	$(I/K)_{t-2}$
				$(I/K)_{t-3}$	$(I/K)_{t-3}$
					$\vdots$
					$(I/K)_1$

Sample: 703 firms; 4966 observations; 1988–2000
Notes: Year dummies included in all models.
Asymptotic standard errors in parentheses.
m1 and m2 are tests for first-order and second-order serial correlation, asymptotically N(0,1). These test the levels residuals for OLS levels, and the first-differenced residuals in all other columns.
GMM results are one-step estimates with heteroskedasticity-consistent standard errors and test statistics.
Sargan is a test of the overidentifying restrictions for the GMM estimators, asymptotically χ^2. P-value is reported. This test uses the minimised value of the corresponding two-step GMM estimators.
All computations done using DPD98 for Gauss; see Arellano and Bond (1998).

3 Multivariate dynamic models with persistent series

3.1 Autoregressive-distributed lag models

The GMM estimators for autoregressive models outlined in the previous section extend in a natural way to autoregressive-distributed lag models of the form

$$y_{it} = \alpha y_{i,t-1} + \beta x_{it} + (\eta_i + v_{it}); \quad i = 1, 2, \ldots, N; \quad t = 2, 3, \ldots, T \quad (3.1)$$

where x_{it} can be a vector of current and lagged values of additional explanatory variables. An attractive feature of this approach is that it does not require models for the x_{it} series to be specified in order to estimate the parameters (α, β).

Different moment conditions will be available depending on what is assumed about the correlation between x_{it} and the two components of the error term. For example, focusing on the case where x_{it} is scalar, we first assume that x_{it} is correlated with the individual effects η_i. The potential to obtain consistent parameter estimates in the presence of this kind of unobserved heterogeneity is another major advantage of panel data. As both x_{it} and y_{it} are correlated with η_i in this case, a transformation like first-differencing is again required to eliminate the individual effects from the transformed equations in order to obtain valid moment conditions.

Maintaining also that the v_{it} disturbances are serially uncorrelated, the x_{it} series may be endogenous in the sense that x_{it} is correlated with v_{it} and earlier shocks, but x_{it} is uncorrelated with $v_{i,t+1}$ and subsequent shocks; predetermined in the sense that x_{it} and v_{it} are also uncorrelated, but x_{it} may still be correlated with $v_{i,t-1}$ and earlier shocks; or strictly exogenous in the sense that x_{it} is uncorrelated with all past, present and future realisations of v_{is}. If x_{it} is assumed to be endogenous then it is treated symmetrically with the dependent variable y_{it}. In this case the lagged values $x_{i,t-2}$, $x_{i,t-3}$ and longer lags (when observed) will be valid instrumental variables in the first-differenced equations for periods $t = 3, 4, \ldots, T$. Maintaining also that the initial conditions y_{i1} are predetermined, the complete set of moment conditions available has the form of (2.4), in which the vector $(y_{i1}, \ldots, y_{i,t-2})$ is replaced by the vector $(y_{i1}, \ldots, y_{i,t-2}, x_{i1}, \ldots, x_{i,t-2})$ in forming each row of the instrument matrix Z_i in (2.3). Computation of the one-step and two-step first-differenced GMM estimators can then proceed just as in the autoregressive case.

If we make the stronger assumption that there is no contemporaneous correlation and the x_{it} series is predetermined, then $x_{i,t-1}$ is additionally available as a valid instrument in the first-differenced equation for period t. In this case $(y_{i1}, \ldots, y_{i,t-2})$ is replaced by the vector $(y_{i1}, \ldots, y_{i,t-2}, x_{i1}, \ldots, x_{i,t-2}, x_{i,t-1})$ to form the instrument matrix Z_i as in (2.3). If we make the much stronger assumption that x_{it} is strictly exogenous, then the complete time series $x_i' = (x_{i1}, x_{i2}, \ldots, x_{iT})$ will be valid instrumental variables in each of the first-differenced equations. In this case $(y_{i1}, \ldots, y_{i,t-2})$ is replaced by the vector $(y_{i1}, \ldots, y_{i,t-2}, x_i')$ to form the instrument matrix.

Whilst the choice between these alternatives may seem a little arbitrary, in most cases these moment conditions will be overidentifying restrictions, so that the validity of a particular assumption may be tested using standard GMM tests of overidentifying restrictions. Difference Sargan tests are useful in this context, as the set of moment conditions specified under a relatively weak assumption (e.g. x_{it} is endogenous) is a strict subset of the set of moment conditions specified under a stronger assumption (e.g. x_{it} is predetermined). Letting S denote the Sargan statistic obtained under the stronger assumption and S' denote the Sargan statistic obtained under the weaker assumption, the simple difference $DS = S - S'$ is asymptotically χ^2, and tests the validity of the additional moment conditions used in the former case; see Arellano and Bond (1991).

Further moment conditions are available if we are willing to assume that x_{it} is uncorrelated with the unobserved individual effects η_i. The basic difference is that now we have valid instrumental variables for the untransformed levels equations. Unfortunately the complete set of moment conditions specified typically cannot be written as a set of orthogonality conditions using the error term in the levels equations $(\eta_i + v_{it})$ alone. Efficient estimation combines the set of moment conditions available for the first-differenced equations, described above, with the additional moment conditions implied for the levels equations under the additional assumption that x_{it} and η_i are uncorrelated.

For example, if x_{it} is endogenous with respect to v_{it} and uncorrelated with η_i, there are a further $T-1$ non-redundant additional moment conditions of the form

$$E\left[x_{i,t-1}(\eta_i + v_{it})\right] = 0 \quad \text{for } i = 1, 2, \ldots, N \text{ and } t = 2, 3, \ldots, T,$$

in addition to the set of moment conditions for the first-differenced equations described previously.[23] This allows the use of $x_{i,t-1}$ as an instrumental variable in the levels equation for period t. If x_{it} is either predetermined or strictly exogenous with respect to v_{it}, there are an additional T non-redundant moment conditions, given the respective sets of moment conditions for the first-differenced equations under these assumptions. In both cases these can be written as[24]

$$\begin{aligned} E\left[x_{it}(\eta_i + v_{it})\right] &= 0 \quad \text{for } i = 1, 2, \ldots, N \text{ and } t = 2, 3, \ldots, T \\ \text{and } E\left[x_{i1}(\eta_i + v_{i2})\right] &= 0 \quad \text{for } i = 1, 2, \ldots, N; \end{aligned}$$

see Arellano and Bond (1991) for details.

Estimation then combines the set of moment conditions specified for the equations in first-differences with these additional moment conditions specified for the equations in levels. One difference from the case where only first-differenced equations are used is that there is no longer a known weight matrix that can be used to construct a one-step GMM estimator that is asymptotically equivalent to the optimal two-step GMM estimator, even in the special case of homoskedastic disturbances. The gain in efficiency from using the optimal two-step weight matrix is thus likely to be greater in this context, at least in large samples. Note also that the Difference Sargan test can be used to test the assumption that x_{it} is uncorrelated with the individual effects.[25]

An intermediate case occurs when we are not willing to assume that the level of the x_{it} variable is uncorrelated with the individual effects, but we are willing to assume that the first-differences Δx_{it} are uncorrelated with η_i. In this case suitably lagged values of Δx_{is} can be used as instrumental variables in the levels equation for period t; see Arellano and Bover (1995) for details.

3.2 Persistent series

This last observation raises the possibility that lagged differences $\Delta y_{i,t-1}$ may also be valid instruments for the levels equations in autoregressive models like (2.1). This turns out to depend on the validity of a stationarity assumption about the initial conditions y_{i1}, which is discussed in detail in Blundell and Bond (1998). Specifically it requires $E\left[\left(y_{i1} - \left(\frac{\eta_i}{1-\alpha}\right)\right)\eta_i\right] = 0$ for $i = 1, \ldots, N$, so that the initial conditions do not deviate systematically from the value $\left(\frac{\eta_i}{1-\alpha}\right)$ which

[23] This representation is not unique. Note that the use of longer lags of x_{is} as instruments for the equations in levels is redundant when all available lags of x_{is} are used as instruments in the first-differenced equations.

[24] Again this representation is not unique.

[25] Arellano (1993) considers an alternative Wald test of this assumption.

the model specifies that the y_{it} series for individual i converges towards from period 2 onwards.[26] This stationary mean assumption implies $E\left[\Delta y_{i2}\eta_i\right]=0$ for $i=1,2,\ldots,N$, which in turn given the autoregressive structure of the model and the mild assumption $E\left[\Delta v_{it}\eta_i\right]=0$ for $i=1,2,\ldots,N$ and $t=3,4,\ldots,T$ implies the additional $T-2$ non-redundant linear moment conditions

$$E\left[\Delta y_{i,t-1}(\eta_i+v_{it})\right]=0 \quad \text{for } i=1,2,\ldots,N \text{ and } t=3,4,\ldots,T, \tag{3.2}$$

in addition to those specified for the first-differenced equations in (2.4). Estimation again requires a combination of equations in first-differences with equations in levels to exploit this complete set of moment conditions. One further observation is that under these additional assumptions, the quadratic moment conditions discussed earlier can also be shown to be redundant when the linear moment conditions (2.4) and (3.2) are all exploited.[27]

The significance of these additional moment conditions in the $AR(1)$ model is that estimation no longer depends exclusively on the first-differenced equations, or more specifically on the use of lagged levels of the series as instruments for the first-difference $\Delta y_{i,t-1}$. The correlation between lagged levels of the series and $\Delta y_{i,t-1}$ becomes weak if either the true value of the parameter α approaches unity, or if the ratio $\text{var}(\eta_i)/\text{var}(v_{it})$ becomes large. In both cases the time series y_{it} becomes highly persistent and lagged levels provide weak instruments for subsequent first-differences. Indeed if we consider the alternative specification

$$y_{it}=\alpha y_{i,t-1}+(\eta_i(1-\alpha)+v_{it}); \qquad i=1,2,\ldots,N; \quad t=2,3,\ldots,T,$$

which approaches a pure random walk as $\alpha\to 1$, then the parameter α is not identified using the moment conditions for the first-differenced equations (2.4) in the case when $\alpha=1$.

More generally the instruments available for the equations in first-differences are likely to be weak when the individual series have near unit root properties. Instrumental variable estimators can be subject to serious finite sample biases when the instruments used are weak.[28] Blundell and Bond (1998) show that this applies to the first-differenced GMM estimator for the $AR(1)$ model.[29] Table 2 reports a selection of their simulation results, for both the first-differenced GMM estimator and the extended GMM estimator which uses the additional moment conditions (3.2). The Within Groups estimator is added here for comparison. Notice that the finite sample bias for the first-differenced GMM estimator in this context is downwards, and in this experiment where the moment conditions (2.4) are valid but only

---

26 In a specification with period-specific intercepts or time dummies, the requirement that the mean of the $(y_{i1},y_{i2},\ldots,y_{iT})$ series is *constant* for each individual can be replaced by the weaker assumption that these means evolve over time in a *common* way for the N individuals in the sample - so that the means are again constant when the series are expressed as deviations from period-specific means.

27 Further moment conditions are available if the v_{it} disturbances are homoskedastic and the initial conditions y_{i1} satisfy covariance stationarity. See, for example, Ahn and Schmidt (1997) and Kruiniger (2000).

28 See, for example, Bound, Jaeger and Baker (1995) and Staiger and Stock (1997) .

29 Blundell, Bond and Windmeijer (2000) suggest that this is also likely to be a concern in multivariate models when the individual series are highly persistent.

Table 2. Simulation results for the $AR(1)$ model

N	α	Within groups	GMM DIF	GMM system
100	0.5	−0.0370	0.4641	0.5100
		(.0697)	(.2674)	(.1330)
	0.8	0.1343	0.4844	0.8101
		(.0726)	(.8224)	(.1618)
	0.9	0.1906	0.2264	0.9405
		(.0725)	(.8264)	(.1564)
500	0.5	−0.0360	0.4887	0.5021
		(.0310)	(.1172)	(.0632)
	0.8	0.1364	0.7386	0.7939
		(.0328)	(.3085)	(.0779)
	0.9	0.1930	0.5978	0.9043
		(.0330)	(.6407)	(.0999)

Notes: The table reports means (standard deviations) from experiments with $T = 4$ and 1000 replications. The model is $y_{it} = \alpha y_{i,t-1} + (\eta_i + v_{it})$, with $\text{var}(\eta_i) = \text{var}(v_{it}) = 1$ and initial conditions drawn from the covariance stationary distribution for y_{i1}. Results are reported for two-step GMM estimators.

Source: Blundell and Bond (1998) and author's calculations.

weakly identify the parameter α, this finite sample bias appears to be in the direction of the Within Groups estimator. By contrast the extended estimator, denoted system GMM, has much smaller finite sample bias and much greater precision when estimating autoregressive parameters using persistent series.

This emphasis on the time series properties of the series might appear surprising given our focus on micro panels, with a large number of cross-section units and a small number of time periods. In this context the asymptotic distribution theory relies on N becoming large with T treated as fixed, and the asymptotic normality of these GMM estimators does not depend on the time series properties of the series. The issue here is that parameters may not be identified using first-differenced GMM estimators when the series are random walks, and more generally identification may be weak when the series are near unit root processes. As Table 2 illustrates, finite sample biases can be severe when the instruments available are weak. Investigating the time series properties of the individual series is therefore to be recommended when using these GMM estimators for dynamic panel data models, and the validity of the additional moment conditions discussed in this section should be given careful consideration if the series being used are found to be highly persistent. An example which illustrates these issues is presented in the next section.

The extended system GMM estimator also extends straightforwardly to autoregressive-distributed lag models like (3.1). In this context Blundell and Bond (2000) show that a sufficient condition for the validity of additional moment conditions like (3.2) is that the (y_{it}, x_{it}) series each satisfy a mean stationarity assumption. However weaker conditions can also suffice. For example, within our sample the first-differences Δy_{it} can be expressed as

$$\Delta y_{it} = \alpha^{t-2}\Delta y_{i2} + \sum_{s=0}^{t-3} \alpha^s \beta \Delta x_{i,t-s} + \sum_{s=0}^{t-3} \alpha^s \Delta v_{i,t-s}$$
$$\text{for } i = 1, 2, \ldots, N \text{ and } t = 3, 4, \ldots, T.$$

Thus if $\alpha < 1$ and the same model has generated the y_{it} series for sufficiently long prior to our sample period for any influence of the *true* initial conditions to have become negligible, then orthogonality between Δy_{it} and η_i would be implied if all the Δx_{it} series are also uncorrelated with the individual effects. Again it should be noted that the validity of these additional moment conditions for the levels equations can be tested, for example using Difference Sargan tests.

3.3 Example: production function

We illustrate the use of GMM estimators which exploit these additional moment restrictions for the equations in levels in a multivariate context by considering the dynamic production function specification estimated in Blundell and Bond (2000).[30] A nice feature of this example is that it illustrates why adopting a dynamic econometric specification may sometimes be necessary for identifying parameters of interest, even when the dynamics themselves are not the principal focus of attention.

Blundell and Bond (2000) consider the Cobb-Douglas production function

$$\begin{aligned} y_{it} &= \beta_n n_{it} + \beta_k k_{it} + \gamma_t + (\eta_i + v_{it} + m_{it}) \\ v_{it} &= \alpha v_{i,t-1} + e_{it} \qquad |\alpha| < 1 \\ e_{it}, m_{it} &\sim MA(0) \end{aligned} \tag{3.3}$$

where y_{it} is log sales of firm i in year t, n_{it} is log employment, k_{it} is log capital stock and γ_t is a year-specific intercept reflecting, for example, a common technology shock. Of the error components, η_i is an unobserved time-invariant firm-specific effect, v_{it} is a possibly autoregressive (productivity) shock and m_{it} reflects serially uncorrelated (measurement) errors. Constant returns to scale would imply $\beta_n + \beta_k = 1$, but this is not necessarily imposed.

Interest is in the consistent estimation of the parameters $(\beta_n, \beta_k, \alpha)$ when the number of firms (N) is large and the number of years (T) is fixed. It is assumed that both employment (n_{it}) and capital (k_{it}) are potentially correlated with the firm-specific effects (η_i), and with both productivity shocks (e_{it}) and measurement errors (m_{it}).

[30] The discussion in this section draws heavily on that in Blundell and Bond (2000) and that in Blundell, Bond and Windmeijer (2000).

In this case there are no valid moment conditions for the static specification (3.3) if the disturbances v_{it} are indeed autoregressive ($\alpha \neq 0$). However this model has a dynamic (common factor) representation

$$\begin{aligned} y_{it} = \beta_n n_{it} - \alpha\beta_n n_{i,t-1} + \beta_k k_{it} - \alpha\beta_k k_{i,t-1} + \alpha y_{i,t-1} \\ +(\gamma_t - \alpha\gamma_{t-1}) + (\eta_i(1-\alpha) + e_{it} + m_{it} - \alpha m_{i,t-1}) \end{aligned} \quad (3.4)$$

or

$$y_{it} = \pi_1 n_{it} + \pi_2 n_{i,t-1} + \pi_3 k_{it} + \pi_4 k_{i,t-1} + \pi_5 y_{i,t-1} + \gamma_t^* + (\eta_i^* + w_{it}) \quad (3.5)$$

subject to two non-linear (common factor) restrictions $\pi_2 = -\pi_1\pi_5$ and $\pi_4 = -\pi_3\pi_5$. Whilst the error term $(v_{it} + m_{it})$ in (3.3) is serially correlated at all lag lengths, the error term w_{it} is serially uncorrelated if there are no measurement errors ($w_{it} = e_{it}$ with $\text{var}\,(m_{it}) = 0$), or $w_{it} \sim MA(1)$ if there are transient measurement errors in some of the series. In either case we can obtain consistent estimates of the unrestricted parameter vector $\pi = (\pi_1, \pi_2, \pi_3, \pi_4, \pi_5)$ using the GMM methods outlined in the previous sections. Given consistent estimates of π and $\text{var}(\pi)$, the common factor restrictions can be (tested and) imposed using minimum distance to obtain consistent estimates of the restricted parameter vector $(\beta_n, \beta_k, \alpha)$.

Blundell and Bond (2000) consider the time series properties of these series and report estimates of this production function using a balanced panel of 509 R&D-performing US manufacturing companies observed for 8 years, 1982-89, similar to that used in Mairesse and Hall (1996). Capital stock and employment are measured at the end of the firm's accounting year, and sales is used as a proxy for output. Further details of the data construction can be found in Mairesse and Hall (1996).

Table 3 reports simple $AR(1)$ specifications for the three series, employment (n_{it}), capital (k_{it}) and sales (y_{it}). All three series are found to be highly persistent, although even using OLS levels estimates none appears to have an exact unit root.[31] For the employment series, both differenced and system GMM estimators suggest an autoregressive coefficient around 0.9, and differenced GMM does not appear to be seriously biased. However for capital and sales, whilst system GMM again suggests an autoregressive coefficient around 0.9, the differenced GMM estimates are found to be significantly lower, and close to the corresponding Within Groups estimates. These downward biases in differenced GMM estimates of the $AR(1)$ models for capital and sales are consistent with the finite sample biases expected in the case of highly persistent series, illustrated in Table 2. Indeed the surprise is that differenced GMM gives reasonable results for the employment series. One difference is that the variance of the firm-specific effects is found to be lower, relative to the variance of transitory shocks, for the employment series. The ratio of these variances is around 1.2 for employment, but 2.2 for capital and 1.7 for sales.

Table 4 presents results for the basic production function, not imposing constant returns to scale, for a range of estimators. Results are reported for both the unrestricted model (3.5) and the restricted model (3.3), where the common factor

[31] See Bond, Nauges and Windmeijer (2002) and Hall and Mairesse (2001) for further discussion of unit root tests in this context.

Table 3. AR(1) specifications for the production series

Labour (n_t)	OLS levels	Within groups	GMM-DIF $t-3$	GMM-SYS $t-3$	GMM-SYS $t-4$
n_{t-1}	0.986	0.723	0.920	0.923	
	(.002)	(.022)	(.062)	(.033)	
m1	4.16	−8.51	−7.62	−8.99	
m2	2.67	0.60	0.44	0.43	
Sargan			.040	.056	
Dif-Sar				.387	
Capital (k_t)					
k_{t-1}	0.987	0.733	0.768	0.925	
	(.002)	(.027)	(.070)	(.021)	
m1	7.72	−6.82	−5.80	−6.51	
m2	2.29	−1.73	−1.73	−1.81	
Sargan			.563	.627	
Dif-Sar				.562	
Sales (y_t)					
y_{t-1}	0.988	0.693	0.775	0.963	0.893
	(.002)	(.025)	(.063)	(.048)	(.063)
m1	5.70	−7.35	−5.95	−7.15	−6.35
m2	0.97	−2.37	−2.46	−2.53	−2.63
Sargan			.040	.025	.092
Dif-Sar				.134	

See notes to Table 1.
Dif-Sar is the Difference Sargan test. P-value is reported.

Source: Blundell and Bond (2000).

restrictions are tested and imposed using minimum distance. The results here are for one-step GMM estimators, with heteroskedasticity-consistent asymptotic standard errors reported.

As expected in the presence of firm-specific effects, OLS levels appears to give an upwards-biased estimate of the coefficient on the lagged dependent variable, whilst Within Groups appears to give a downwards-biased estimate of this coefficient. Note that even using OLS, the hypothesis that $\alpha = 1$ is rejected, and even using Within Groups the hypothesis that $\alpha = 0$ is rejected. Although the pattern of signs on current and lagged regressors in the unrestricted models are consistent with the $AR(1)$ error-component specification, the common factor restrictions are rejected for both these estimators. They also reject constant returns to scale.[32]

[32] The table reports p-values from minimum distance tests of the common factor restrictions and Wald tests of the constant returns to scale restrictions.

Table 4. Production function estimates

	OLS levels	Within groups	GMM DIF $t-2$	GMM DIF $t-3$	GMM SYS $t-2$	GMM SYS $t-3$
n_t	0.479	0.488	0.513	0.499	0.629	0.472
	(0.029)	(0.030)	(0.089)	(0.101)	(0.106)	(0.112)
n_{t-1}	−0.423	−0.023	0.073	−0.147	−0.092	−0.278
	(0.031)	(0.034)	(0.093)	(0.113)	(0.108)	(0.120)
k_t	0.235	0.177	0.132	0.194	0.361	0.398
	(0.035)	(0.034)	(0.118)	(0.154)	(0.129)	(0.152)
k_{t-1}	−0.212	−0.131	−0.207	−0.105	−0.326	−0.209
	(0.035)	(0.025)	(0.095)	(0.110)	(0.104)	(0.119)
y_{t-1}	0.922	0.404	0.326	0.426	0.462	0.602
	(0.011)	(0.029)	(0.052)	(0.079)	(0.051)	(0.098)
m1	−2.60	−8.89	−6.21	−4.84	−8.14	−6.53
m2	−2.06	−1.09	−1.36	−0.69	−0.59	−0.35
Sargan	–	–	.001	.073	.000	.032
Dif-Sar	–	–	–	–	.001	.102
β_n	0.538	0.488	0.583	0.515	0.773	0.479
	(0.025)	(0.030)	(0.085)	(0.099)	(0.093)	(0.098)
β_k	0.266	0.199	0.062	0.225	0.231	0.492
	(0.032)	(0.033)	(0.079)	(0.126)	(0.075)	(0.074)
α	0.964	0.512	0.377	0.448	0.509	0.565
	(0.006)	(0.022)	(0.049)	(0.073)	(0.048)	(0.078)
Comfac	.000	.000	.014	.711	.012	.772
CRS	.000	.000	.000	.006	.922	.641

See notes to Tables 1 and 3.
Comfac is a minimum distance test of the non-linear common factor restrictions imposed in the restricted models. P-values are reported.
CRS is a Wald test of the constant returns to scale hypothesis $\beta_n + \beta_k = 1$ in the restricted models. P-values are reported.

Source: Blundell and Bond (2000).

The validity of lagged levels dated $t-2$ as instruments in the first-differenced equations is clearly rejected by the Sargan test of overidentifying restrictions.[33] This is consistent with the presence of measurement errors. Instruments dated $t-3$ (and earlier) are not rejected, and the test of common factor restrictions is easily passed in these first-differenced GMM results. However the estimated coefficient on the lagged dependent variable is barely higher than the Within Groups estimate, suggesting the possibility of serious finite sample bias here. Indeed the differenced GMM parameter estimates are all very close to the Within Groups results. Serious

[33] Again the Sargan tests reported here are based on the minimised values of the associated two-step GMM estimators.

finite sample bias associated with weak instruments may well be suspected in this context, given the time series properties of these series discussed above. The estimate of β_k is low and statistically weak, and the constant returns to scale restriction is rejected.

The validity of lagged levels dated $t-3$ (and earlier) as instruments in the first-differenced equations, combined with lagged first-differences dated $t-2$ as instruments in the levels equations, appears to be marginal in the system GMM estimator. However the Difference Sargan statistic that specifically tests the additional moment conditions used in the levels equations accepts their validity at the 10% level. The system GMM parameter estimates appear to be reasonable. The estimated coefficient on the lagged dependent variable is higher than the Within Groups estimate, but well below the OLS levels estimate. The common factor restrictions are easily accepted, and the estimate of β_k is both higher and better determined than the differenced GMM estimate. The constant returns to scale restriction is easily accepted in the system GMM results.

Blundell and Bond (2000) explore this data in more detail and conclude that the system GMM estimates in the final column of Table 4 are their preferred results. They also report that when constant returns to scale is imposed on the production function - it is not rejected in the preferred system GMM results - then the results obtained using the first-differenced GMM estimator become more similar to the system GMM estimates.

4 Summary and conclusions

This paper has reviewed the use of Generalised Method of Moments estimators in the context of single equation, autoregressive-distributed lag models estimated from micro panels with a large number of cross-section units observed for a small number of time periods. These methods are particularly useful when the model of interest contains endogenous or predetermined explanatory variables, but the processes generating these series are not completely specified.

These GMM estimators can be used to obtain consistent parameter estimates in a wide range of microeconometric applications. However we have emphasised that they may be subject to large finite sample biases when the instruments available are weak, and this is particularly likely to be a problem when using the basic first-differenced estimators with series that are highly persistent. Careful investigation of the time series properties of the individual series, and comparison of the consistent GMM estimators to simpler estimators like OLS levels and Within Groups, which are likely to be biased in opposite directions in the context of coefficients on lagged dependent variables in short T panels, can help in detecting and avoiding these biases in applied research.

References

Ahn S C, Schmidt P (1995) Efficient estimation of models for dynamic panel data. Journal of Econometrics 68: 5–28

Ahn S C, Schmidt P (1997) Efficient estimation of dynamic panel data models: alternative assumptions and simplified estimation. Journal of Econometrics 76: 309–321

Anderson T W, Hsiao C (1981) Estimation of dynamic models with error components. Journal of the American Statistical Association 76: 598–606

Anderson T W, Hsiao C (1982) Formulation and estimation of dynamic models using panel data. Journal of Econometrics 18: 47–82

Arellano, M (1993) On the testing of correlated effects with panel data. Journal of Econometrics 59: 87–97

Arellano M, Bond S R (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 R (1998) Dynamic panel data estimation using DPD98 for GAUSS. Mimeo, Institute for Fiscal Studies, London (http://www.ifs.org.uk/staff/steve_b.shtml)

Arellano M, Bover O (1995) Another look at the instrumental-variable estimation of error-components models. Journal of Econometrics 68: 29–52

Arellano M, Honore B (2001) Panel data models: some recent developments. In: Heckman J J, Leamer E E (eds) Handbook of econometrics, Vol 5. North-Holland, Amsterdam

Binder M., Hsiao C, Pesaran M H (2000) Estimation and inference in short panel vector autoregressions with unit roots and cointegration. University of Cambridge DAE Working Paper No. 0003

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

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

Blundell R W, Bond S R, Windmeijer F (2000) Estimation in dynamic panel data models: improving on the performance of the standard GMM estimator. In: Baltagi B (ed) Advances in econometrics, Vol 15. Nonstationary panels, panel cointegration, and dynamic panels. JAI Elsevier, Amsterdam

Blundell R W, Smith R J (1991) Initial conditions and efficient estimation in dynamic panel data models. Annales d'Economie et de Statistique 20/21: 109–123

Bond S R, Klemm A, Newton-Smith R, Syed M H, Vlieghe G (2002) The roles of expected profitability, Tobin's q and cash flow in econometric models of company investment. Mimeo, Institute for Fiscal Studies, London

Bond S R , Nauges C, Windmeijer F (2002) Unit roots and identification in autoregressive panel data models: a comparison of alternative tests. Mimeo, Institute for Fiscal Studies, London

Bond S R, Windmeijer F (2002) Finite sample inference for GMM estimators in linear panel data models. Cemmap Working Paper Series No. CWP04/02, Institute for Fiscal Studies, London

Bound J, Jaeger D A, Baker R M (1995) Problems with instrumental variables estimation when the correlation between the instruments and the endogenous explanatory variable is weak. Journal of the American Statistical Association 90: 443–450

Bowsher C (2000) On testing overidentifying restrictions in dynamic panel data models. Nuffield College Discussion Paper in Economics, no. 2000-W28

Hall B H, Mairesse J (2001) Testing for unit roots in panel data: an exploration using real and simulated data. Mimeo, UC Berkeley

Hansen L P (1982) Large sample properties of generalised method of moment estimators. Econometrica 50: 1029–1054

Holtz-Eakin D, Newey W, Rosen H S (1988) Estimating vector autoregressions with panel data. Econometrica 56: 1371–1396

Hsiao C (1986) Analysis of panel data. Cambridge University Press, Cambridge

Kruiniger H (2000) GMM estimation of dynamic panel data models with persistent data. QMW, University of London, Dept. of Economics Working Paper Series No. 428

Mairesse J, Hall B H (1996) Estimating the productivity of research and development in French and US manufacturing firms: an exploration of simultaneity issues with GMM methods. In: Wagner K, Van

Ark B (eds) International productivity differences and their explanations, pp 285–315. Elsevier, Amsterdam
Nickell S J (1981) Biases in dynamic models with fixed effects. Econometrica 49: 1417–1426
Nickell S J (1986) Dynamic models of labor demand. In: Ashenfelter O, Layard R (eds) Handbook of labor economics, Vol 1. North-Holland, Amsterdam
Sargan J D (1958) The estimation of economic relationships using instrumental variables. Econometrica 26: 329–338
Sargan J D (1988) Testing for misspecification after estimation using instrumental variables. In: Maasoumi E (ed) Contributions to econometrics: John Denis Sargan, Vol I. Cambridge University Press, Cambridge
Staiger D, Stock J H (1997) Instrumental variables regression with weak instruments. Econometrica 65: 557–586
Verbeek M (1996) Pseudo panel data. In: Matyas L, Sevestre P (eds) The econometrics of panel data, 2nd edn. Kluwer, Amsterdam
Windmeijer F (2000) A finite sample correction for the variance of linear two-step GMM estimators. Institute for Fiscal Studies Working Paper Series No. W00/19, London