Some Tests of Specification for Panel Data: Monte Carlo Evidence and an Application to Employment Equations

Author(s): Manuel Arellano and Stephen Bond

Source: *The Review of Economic Studies*, Apr., 1991, Vol. 58, No. 2 (Apr., 1991), pp. 277-297

Published by: Oxford University Press

Stable URL: https://www.jstor.org/stable/2297968

Review of Economic Studies (1991) **58**, 277-297 0034-6527/91/00180277$02.00

Some Tests of Specification for Panel Data: Monte Carlo Evidence and an Application to Employment Equations

MANUEL ARELLANO
London School of Economics

and

STEPHEN BOND
University of Oxford

First version received May 1988; *final version accepted July* 1990 (*Eds.*)

This paper presents specification tests that are applicable after estimating a dynamic model from panel data by the generalized method of moments (GMM), and studies the practical performance of these procedures using both generated and real data. Our GMM estimator optimally exploits all the linear moment restrictions that follow from the assumption of no serial correlation in the errors, in an equation which contains individual effects, lagged dependent variables and no strictly exogenous variables. We propose a test of serial correlation based on the GMM residuals and compare this with Sargan tests of over-identifying restrictions and Hausman specification tests.

1. INTRODUCTION

The purpose of this paper is to present specification tests that are applicable after estimating a dynamic model from panel data by the generalized method of moments (GMM) and to study the practical performance of these procedures using both generated and real data.

Previous work concerning dynamic equations from panel data (e.g. Chamberlain (1984), Bhargava and Sargan (1983)) has emphasized the case where the model with an arbitrary intertemporal covariance matrix of the errors is identified. The fundamental identification condition for this model is the strict exogeneity of some of the explanatory variables (or the availability of strictly exogenous instrumental variables) conditional on the unobservable individual effects. In practice, this allows one to use past, present and future values of the strictly exogenous variables to construct instruments for the lagged dependent variables and other non-exogenous variables once the permanent effects have been differenced out. Bhargava and Sargan (1983) and Arellano (1990) considered estimation and inference imposing restrictions on the autocovariances, but the assumption that the model with unrestricted covariance matrix is identified was never removed.

However, sometimes one is less willing to assume the strict exogeneity of an explanatory variable than to restrict the serial correlation structure of the errors, in which case different identification arrangements become available. Uncorrelated errors arise in a number of environments. These include rational expectations models where the disturbance is a surprise term, error-correction models and vector autoregressions. Moreover, if there are a priori reasons to expect autoregressive errors in a regression model, these can be represented as a dynamic regression with non-linear common factor restrictions and uncorrelated disturbances (e.g. Sargan (1980)). In these cases and also in models with moving-average errors, lagged values of the dependent variable itself become valid instruments in the differenced equations corresponding to later periods. Simple estimators of this type were first proposed by Anderson and Hsiao (1981, 1982). Griliches and Hausman (1986) have developed estimators for errors-in-variables models whose identification relies on assumptions of lack of (or limited) serial correlation in the measurement errors. Holtz-Eakin, Newey and Rosen (1988) have also considered estimators of this type for vector autoregressions which are similar to the ones we employ in this paper.

An estimator that uses lags as instruments under the assumption of white noise errors would lose its consistency if in fact the errors were serially correlated. It is therefore essential to satisfy oneself that this is not the case by reporting test statistics of the validity of the instrumental variables (i.e. tests of lack of serial correlation) together with the parameter estimates. In this paper we consider three such tests: a direct test on the second-order residual serial correlation coefficient, a Sargan test of over-identifying restrictions and a Hausman specification test. The operating characteristics of these tests are different as well as their number of degrees of freedom. In addition, depending on alternative auxiliary distributional assumptions concerning stationarity and heterogeneity, different forms of each of the tests are available. These alternative versions of a given test are asymptotically equivalent under the less general set of auxiliary assumptions but they still may perform quite differently in finite samples. We have therefore produced a number of Monte Carlo experiments to investigate the relative performance of the various tests. Finally, as an empirical illustration we report some estimated employment equations using the Datastream panel of quoted U.K. companies.

The paper is organized as follows. Section 2 presents the model and the estimators. For a fixed number of time periods in the sample, the model specifies a finite number of moment restrictions and therefore an asymptotically efficient GMM estimator is readily available. The discussion is kept as simple as possible by concentrating initially on a first-order autoregression with a fixed effect. Exogenous variables and unbalanced panel considerations are subsequently introduced. Section 3 presents the various tests of serial correlation and their asymptotic distributions. Section 4 reports the simulation results. Section 5 contains the application to employment equations and Section 6 concludes.

2. ESTIMATION

The simplest model without strictly exogenous variables is an autoregressive specification of the form

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

We assume that a random sample of N individual time series $(y_{i1}, \ldots, y_{iT})$ is available. T is small and N is large. The v_{it} are assumed to have finite moments and in particular $E(v_{it}) = E(v_{it}v_{is}) = 0$ for $t \neq s$. That is, we assume lack of serial correlation but not necessarily independence over time. With these assumptions, values of y lagged two periods or more are valid instruments in the equations in first differences. Namely, for

$T \geqq 3$ the model implies the following $m=(T-2)(T-1)/2$ linear moment restrictions

$$E[(\bar{y}_{it}-\alpha\bar{y}_{i(t-1)})y_{i(t-j)}]=0 \qquad (j=2,\ldots,(t-1);\ t=3,\ldots,T) \tag{2}$$

where for simplicity $\bar{y}_{it}=y_{it}-y_{i(t-1)}$. We wish to obtain the optimal estimator of α as $N\rightarrow\infty$ for fixed T on the basis of these moment restrictions alone. That is, in the absence of any other knowledge concerning initial conditions or the distributions of the v_{it} and the η_i. Note that our assumptions also imply quadratic moment restrictions, for example $E(\bar{v}_{it}\bar{v}_{i(t-2)})=0$, which however we shall not exploit in order to avoid iterative procedures.

This estimation problem is an example of those analyzed by Hansen (1982) and White (1982), and an optimal GMM or two-stage instrumental variables estimator should be available. The moment equations in (2) can be conveniently written in vector form as $E(Z_i'\bar{v}_i)=0$ where $\bar{v}_i=(\bar{v}_{i3}\cdots\bar{v}_{iT})'$ and Z_i is a $(T-2)\times m$ block diagonal matrix whose sth block is given by $(y_{i1}\cdots y_{is})$.[1]

The GMM estimator $\hat{\alpha}$ is based on the sample moments $N^{-1}\sum_{i=1}^{N}Z_i'\bar{v}_i=N^{-1}Z'\bar{v}$ where $\bar{v}=\bar{y}-\alpha\bar{y}_{-1}=(\bar{v}_1',\ldots,\bar{v}_N')'$ is a $N(T-2)\times 1$ vector and $Z=(Z_1',\ldots,Z_N')'$ is a $N(T-2)\times m$ matrix. $\hat{\alpha}$ is given by

$$\hat{\alpha}=\text{argmin}_{\alpha}\ (\bar{v}'Z)A_N(Z'\bar{v})=\frac{\bar{y}_{-1}'ZA_NZ'\bar{y}}{\bar{y}_{-1}'ZA_NZ'\bar{y}_{-1}}. \tag{3}$$

Multivariate standard CLT implies that $\bar{V}_N^{-1/2}N^{-1/2}Z'\bar{v}$ is asymptotically standard normal where $\bar{V}_N=N^{-1}\sum_i E(Z_i'\bar{v}_i\bar{v}_i'Z_i)$ is the average covariance matrix of $Z_i'\bar{v}_i$. Under our assumptions, $\bar{V}_N$ can be replaced by $\hat{V}_N=N^{-1}\sum_i Z_i'\hat{\bar{v}}_i\hat{\bar{v}}_i'Z_i$ where the $\bar{v}_i$ are residuals from a preliminary consistent estimator $\hat{\alpha}_1$. The one-step estimator $\hat{\alpha}_1$ is obtained by setting $A_N=(N^{-1}\sum_i Z_i'HZ_i)^{-1}$ where H is a $(T-2)$ square matrix which has twos in the main diagonal, minus ones in the first subdiagonals and zeroes otherwise. A consistent estimate of avar $(\hat{\alpha})$ for arbitrary A_N is given by

$$\text{av}\hat{\text{a}}\text{r}\,(\hat{\alpha})=N\frac{\bar{y}_{-1}'ZA_N\hat{V}_NA_NZ'\bar{y}_{-1}}{(\bar{y}_{-1}'ZA_NZ'\bar{y}_{-1})^2}. \tag{4}$$

The optimal choice for A_N is $\hat{V}_N^{-1}$ (e.g. see Hansen (1982)) which produces a two-step estimator $\hat{\alpha}_2$.[2] $\hat{\alpha}_1$ and $\hat{\alpha}_2$ will be asymptotically equivalent if the v_{it} are independent and homoskedastic both across units and over time.

It is useful to relate these estimators to the Anderson–Hsiao (AH) estimator which is commonly used in practice. Anderson and Hsiao (1981) proposed to estimate α by regressing $\bar{y}$ on $\bar{y}_{-1}$ using either $\bar{y}_{-2}$ or y_{-2} as instruments. Since both $\bar{y}_{-2}$ and y_{-2} are linear combinations of Z the resulting estimators will be inefficient. Note that under stationarity, namely when $E(y_{it}y_{i(t-k)})=c_{ik}$ for all t, the estimator that uses $Z_i^+=\text{diag}\,(y_{it})$ $(t=1,\ldots,T-2)$ is asymptotically equivalent to the one based on the stacked vector y_{-2}, whose computation is much simpler (since A_N becomes irrelevant). However neither of the two is asymptotically equivalent to $\hat{\alpha}_1$ or $\hat{\alpha}_2$, not even under stationarity.

The extension of the previous results to the case where a limited amount of serial correlation is allowed in the v_{it} is straightforward. Suppose that v_{it} is MA (q) in the

1. In this paper we represent this type of matrix by $Z_i=\text{diag}\,(y_{i1},\ldots,y_{is})$, $(s=1,\ldots,T-2)$.

2. An alternative choice of A_N is $(N^{-1}\sum_i Z_i'\hat{\Omega}Z_i)^{-1}$ with $\hat{\Omega}=N^{-1}\sum_i\hat{\bar{v}}_i\hat{\bar{v}}_i'$. The resulting estimator does not depend on the data fourth-order moments and is asymptotically equivalent to $\hat{\alpha}_2$ provided the v_{it} are serially independent. Note that in this case $E(Z_i'\bar{v}_i\bar{v}_i'Z_i)=E(Z_i'\Omega_iZ_i)$ and $\lim_{N\rightarrow\infty}N^{-1}\sum_i E[Z_i'(\Omega_i-\bar{\Omega}_N)Z_i]=0$ (see White (1982), p. 492).

sense that $E(v_{it}v_{i(t-k)}) \neq 0$ for $k \leqq q$ and zero otherwise. In this case α is just identified with $T=q+3$ and there are $m_q=(T-q-2)(T-q-1)/2$ restrictions available.

Models with exogenous variables

We now turn to consider an extended version of equation (1) where $(k-1)$ independent explanatory variables have been included

$$y_{it} = \alpha y_{i(t-1)} + \beta' x^*_{it} + \eta_i + v_{it} = \delta' x_{it} + \eta_i + v_{it} \tag{5}$$

where $x_{it} = (y_{i(t-1)}\, x^{*\prime}_{it})'$ is $k \times 1$ and the v_{it} are not serially correlated. Suppose initially that the x^*_{it} are all correlated with η_i. In this context the form of the optimal matrix of instruments depends on whether the x^*_{it} are predetermined or strictly exogenous variables. If the x^*_{it} are predetermined, in the sense that $E(x^*_{it}v_{is}) \neq 0$ for $s<t$ and zero otherwise, then only $x^*_{i1}, \ldots, x^*_{i(s-1)}$ are valid instruments in the differenced equation for period s so that the optimal Z_i is a $(T-2)\times(T-2)[(k-1)(T+1)+(T-1)]/2$ matrix of the form $Z_i = \text{diag}\,(y_{i1} \cdots y_{is} x^{*\prime}_{i1} \cdots x^{*\prime}_{i(s+1)})$, $(s=1,\ldots,T-2)$. On the other hand, if the x^*_{it} are strictly exogenous, i.e. $E(x^*_{it}v_{is})=0$ for all t, s, then all the x^*'s are valid instruments for all the equations and Z_i takes the form $Z_i = \text{diag}\,(y_{i1} \cdots y_{is}\, x^{*\prime}_{i1} \cdots x^{*\prime}_{iT})$, $(s=1,\ldots,T-2)$. Clearly, x^*_{it} may also include a combination of both predetermined and strictly exogenous variables. In either case, the form of the GMM estimator of the $k\times 1$ coefficient vector δ is

$$\hat{\delta} = (\bar{X}' Z A_N Z' \bar{X})^{-1} \bar{X}' Z A_N Z' \bar{y} \tag{6}$$

where $\bar{X}$ is a stacked $(T-2)N \times k$ matrix of observations on $\bar{x}_{it}$, and $\bar{y}$ and Z are as above for the appropriate choice of Z_i. As before, alternative choices of A_N will produce one-step or two-step estimators.[3]

Turning now to the case where x^*_{it} can be partitioned into $(x^*_{1it} x^*_{2it})$ and x^*_{1it} is uncorrelated with η_i, additional moment restrictions exploiting this lack of correlation in the levels equations become available. For example, if x^*_{1it} is predetermined and letting $u_{it} = \eta_i + v_{it}$, we have T extra restrictions. Namely, $E(u_{i2}x^*_{1i1})=0$ and $E(u_{it}x^*_{1it})=0$, $(t=2,\ldots,T)$. Note that all remaining restrictions from the levels equations are redundant given those previously exploited for the equations in first differences. Define $u_i=(u_{i2}\cdots u_{iT})'$, let v^+_i be the $[(T-2)+(T-1)]\times 1$ vector $v^+_i=(\bar{v}'_i u'_i)'$ and let $v^+=(v^{+\prime}_1 \cdots v^{+\prime}_N)' = y^+ - X^+\delta$. The optimal matrix of instruments Z^+_i is block diagonal and consists of two blocks: Z_i which is as in the predetermined x^*_{it} case above (assuming that x^*_{2it} is also predetermined), and Z^a_i which is itself block diagonal with $(x^{*\prime}_{1i1} x^{*\prime}_{1i2})$ in the first block and $x^{*\prime}_{1is}$, $s=3,\ldots,T$ in the remaining blocks. The two-step estimator is of the same form as (6) with X^+, y^+ and Z^+ replacing $\bar{X}, \bar{y}$ and Z respectively, and $A_N=[N^{-1}\sum_i Z^+_i \hat{v}^+_i \hat{v}^{+\prime}_i Z^+_i]^{-1}$. On the other hand, if x^*_{1it} is strictly exogenous, the observations for all periods become valid instruments in the levels equations. However, given those previously exploited in first differences we only have T extra restrictions which in this case can be conveniently expressed as $E(T^{-1}\sum_{s=1}^{T} u_{is}x^*_{1it})=0$ $(t=1,\ldots,T)$. Thus, the two-step estimator would just combine the $(T-1)$ first difference equations and the average level equation.[4]

3. Note that if $E(v_i v'_i)$ is unrestricted (i.e. v_{it} is $MA(q)$ with $T<q+3$) but x^*_{it} is strictly exogenous, the model is identified with $Z_i = \text{diag}\,(x^{*\prime}_{il} \cdots x^{*\prime}_{iT})$, $(s=1,\ldots,T-2)$, in which case the two-step estimator coincides with the generalized three-stage least squares estimator proposed by Chamberlain (1984).

4. Note that when x^*_{1it} variables are available it may be possible to identify and estimate coefficients for time-invariant variables on the lines suggested by Hausman and Taylor (1981), Bhargava and Sargan (1983) and Amemiya and MaCurdy (1986).

Models from unbalanced panel data

By unbalanced panel data we refer to a sample in which consecutive observations on individual units are available, but the number of time periods available may vary from unit to unit as well as the historical points to which the observations correspond. This type of sample is very common particularly with firm data which is the context of the application reported below. Aside from often allowing one to exploit a much larger sample or to pool more than one panel, the use of unbalanced panels may lessen the impact of self selection of firms in the sample. In fact nothing fundamental changes in the econometric methods provided a minimal number of continuous time periods are available on each unit, and one assumes that if period-specific parameters are present the number of observations on these periods tend to infinity. Of course, the essential assumption is that the observations in the initial cross-section are independently distributed and that subsequent additions and deletions take place at random (see Hsiao (1986), Chapter 8).

The previous notation can accommodate unbalanced panels with minor changes. We now have T_i time-series observations on the ith unit, and there are N individual units in the sample. The matrices $\bar{X}$ and Z, and the vectors $\bar{y}$ and $\bar{v}$ are made of N row-blocks, the ith block containing (T_i-2) rows. Note that now the number of non-zero columns in each Z_i may vary across units. For example, in the first-order autoregressive specification above, the number of columns in Z_i is $p=(\tau-2)(\tau-1)/2$ where τ is the total number of periods on which observations are available for some individuals in the sample, and $Z_i = \text{diag}\,(y_{i1},\ldots,y_{is})$, $(s=1,\ldots,\tau-2)$, only if τ observations are available on i. For individuals with $T_i<\tau$, the rows of $\text{diag}\,(y_{i1},\ldots,y_{is})$ corresponding to the missing equations are deleted and the missing values of y_{it} in the remaining rows are replaced by zeroes. The two-step GMM estimator of α for this choice of instruments[5] is the same as in (3) using $A_N=(N^{-1}\sum_i Z_i'\hat{\bar{v}}_i\hat{\bar{v}}_i'Z_i)^{-1}$ where Z_i is $(T_i-2)\times p$ and $\hat{\bar{v}}_i$ is $(T_i-2)\times 1$.

3. TESTING THE SPECIFICATION

In order to keep the notation simple we now drop the bars from variables in first differences, so that the first-difference equation for the unbalanced panel is now

$$\underset{n\times 1}{y} = \underset{n\times k}{X}\ \underset{k\times 1}{\delta} + \underset{n\times 1}{v} \tag{7}$$

where $n=\sum_i (T_i-2)$. We also assume that the x_{it}^* are all potentially correlated with η_i. The $n\times 1$ vector of residuals is given by

$$\hat{v} = y - X\hat{\delta} = v - X(\hat{\delta}-\delta)$$

where $\hat{\delta}$ can be any estimator of the form (6) for a particular choice of Z and A_N. Let $\hat{v}_{-2}$ be the vector of residuals lagged twice, of order $q=\sum_i (T_i-4)$ and let v_* be a $q\times 1$ vector of trimmed v to match v_{-2} and similarly for X_*. Since the v_{it} are first differences of serially uncorrelated errors, $E(v_{it}v_{i(t-1)})$ need not be zero, but the consistency of the GMM estimators above hinges heavily upon the assumption that $E(v_{it}v_{i(t-2)})=0$. In an

5. This is the optimal choice amongst the estimators that can be obtained by stacking the equations for all periods and individuals. An alternative estimator would minimize the sum of the GMM criteria for each of the balanced sub-panels in the sample. Although the latter is strictly more efficient when the number of units in all sub-samples tend to infinity, it may have a poorer finite-sample performance when the various sub-sample sizes are not sufficiently large.

unbalanced panel $(\tau-4)$ such covariances can be estimated in total, in principle with varying number of sample observations to estimate each of the covariances. Provided one assumes that all sub-samples tend to infinity, a $(\tau-4)$ degrees of freedom test can be constructed of the hypothesis that the second-order autocovariances for all periods in the sample are zero. However, a considerably simpler procedure will look at the average covariances $\phi_i = v'_{i(-2)}v_{i*}$. These averages are independent random variables across units with zero mean under the null although with unequal variances in general. So a straightforward one degree of freedom test statistic can be constructed to test whether $E(\phi_i)$ is zero or not.

The test statistic for second-order serial correlation based on residuals from the first-difference equation takes the form

$$m_2 = \frac{\hat{v}'_{-2}\hat{v}_*}{\hat{v}^{1/2}} \,\tilde{a}\, N(0,1) \tag{8}$$

under $E(v_{it}v_{i(t-2)})=0$, where $\hat{v}$ is given by

$$\hat{v} = \textstyle\sum_{i=1}^N v'_{i(-2)}\hat{v}_{i*}v'_{i*}\hat{v}_{i(-2)} - 2\hat{v}'_{-2}X_*(X'ZA_NZ'X)^{-1}X'ZA_N(\sum_{i=1}^N Z'_i\hat{v}_i\hat{v}'_{i*}\hat{v}_{i(-2)}) + \hat{v}'_{-2}X_* \,\text{av}\hat{\text{a}}\text{r}\,(\hat{\delta})X'_*\hat{v}_{-2}. \tag{9}$$

Note that m_2 is only defined if min $T_i \geqq 5$. A proof of the asymptotic normality result is sketched in the Appendix.

It is interesting to notice that the m_2 criterion is rather flexible, in that it can be defined in terms of any consistent GMM estimator, not necessarily in terms of efficient estimators, either in the sense of using optimal Z or A_N or both. However, the asymptotic power of the m_2 test will depend on the efficiency of the estimators used.

The m_2 statistic tests for lack of second-order serial correlation in the first-difference residuals. This will certainly be the case if the errors in the model in levels are not serially correlated, but also if the errors in levels follow a random-walk process. One may attempt to discriminate between the two situations by calculating an m_1 statistic, on the same lines as m_2, to test for lack of first-order serial correlation in the differenced residuals. Alternatively, notice that if the errors in levels follow a random walk, then both OLS and GMM estimates in the first-difference model are consistent which suggests a Hausman test based on the difference between the two estimators.

We now turn to consider two other tests of specification which are applicable in the same context. One is a Sargan test of over-identifying restrictions (cf. Sargan (1958, 1988), Hansen (1982)) given by

$$s = \hat{v}'Z(\textstyle\sum_{i=1}^N Z'_i\hat{v}_i\hat{v}'_iZ_i)^{-1}Z'\hat{v} \,\tilde{a}\, \chi^2_{p-k} \tag{10}$$

where $\hat{v} = y - X\hat{\delta}$, and $\hat{\delta}$ is a two-step estimator of δ for a given Z. Notice that Z need not be the optimal set of instruments; here p just refers to the number of columns in Z provided $p>k$. Also notice that while we are able to produce a version of the serial correlation test based upon a one-step estimator of δ which remained asymptotically normal on the null under the more general distributional assumptions, no robust chi-square Sargan test based on one-step estimates is available. Under the null a statistic of the form

$$s_1 = \frac{1}{\hat{\sigma}^2}\tilde{v}'Z(\textstyle\sum_{i=1}^N Z'_iH_iZ_i)^{-1}Z'\tilde{v}$$

where $\tilde{v}$ are one-step residuals, will only have a limiting chi-square distribution if the errors are indeed i.i.d. over time and individuals. In general, the asymptotic distribution of s_1 is a quadratic form in standard normal variables. Critical values can still be calculated by numerical integration but this clearly leads to a burdensome test procedure.

On the other hand, there may be circumstances where the serial correlation test is not defined while the Sargan test can still be computed. As a simple example, take the first-order autoregressive equation at the beginning of Section 2 with $T=4$; in this case the Sargan statistic tests two linear combinations of the three moment restrictions available, namely $E(\bar{v}_{i3}y_{i1})=E(\bar{v}_{i4}y_{i1})=E(\bar{v}_{i4}y_{i2})=0$, but no differenced residuals two periods apart are available to construct an m_2 test.

A further possibility is to use Sargan difference tests to discriminate between nested hypotheses concerning serial correlation in a sequential way. For example, let Z_I be a $n\times p_I$ matrix containing the columns of Z which remain valid instrumental variables when the errors in levels are first-order moving average, and let $\hat{\delta}_I$ be a two-step estimator of δ based on Z_I with associated residuals $\hat{v}_I$, then

$$s_I=\hat{v}_I Z_I\left(\sum_{i=1}^{N} Z'_{Ii}\hat{v}_{Ii}\hat{v}'_{Ii}Z_{Ii}\right)^{-1}Z'_I\hat{v}_I \mathrel{\tilde{a}} \chi^2_{p_I-k}$$

if the errors in levels are MA (0) or MA (1). In addition

$$ds=s-s_I \mathrel{\tilde{a}} \chi^2_{p-p_I} \tag{11}$$

if the errors in levels are not serially correlated. Moreover ds is asymptotically independent of s_I (see Appendix).

A closely related alternative is to construct a Hausman test based on the difference $(\hat{\delta}_I-\hat{\delta})$ (cf. Hausman (1978) and Hausman and Taylor (1981)). This type of test has been proposed by Griliches and Hausman (1986) in the context of moving-average measurement errors. A test is based on the statistic

$$h=(\hat{\delta}_I-\hat{\delta})'[\widehat{\text{avar}}(\hat{\delta}_I)-\widehat{\text{avar}}(\hat{\delta})]^{-}(\hat{\delta}_I-\hat{\delta}) \mathrel{\tilde{a}} \chi^2_r \tag{12}$$

where $r=\text{rank avar}(\hat{\delta}_I-\hat{\delta})$ and $(\ )^{-}$ indicates a generalized inverse. The value of r will depend on the number of columns of X which are maintained to be strictly exogenous. In particular, if the only non-exogenous variable is the lagged dependent variable then $r=1$.

4. EXPERIMENTAL EVIDENCE

A limited simulation was carried out to study the performance of the estimation and testing procedures discussed above in samples of a size likely to be encountered in practice. In all the experiments the dependent variable y_{it} was generated from a model of the form

$$y_{it}=\alpha y_{i(t-1)}+\beta x_{it}+\eta_i+v_{it}, \qquad (i=1,\dots,N;\ t=1,\dots,T+10)$$
$$v_{it}=\sigma_{it}(\xi_{it}+\phi\xi_{i(t-1)}), \qquad \sigma^2_{it}=\theta_0+\theta_1 x^2_{it} \tag{13}$$

where $\eta_i\sim$ i.i.d. $N(0,\sigma^2_\eta)$, $\xi_{it}\sim$ i.i.d. $N(0,1)$ and $y_{i0}=0$. The first ten cross-sections were discarded so that the actual samples contain NT observations.

With regard to x_{it}, we considered the following generating equation

$$x_{it}=\rho x_{i(t-1)}+\varepsilon_{it} \tag{14}$$

with $\varepsilon_{it}\sim$ i.i.d. $N(0,\sigma^2_\varepsilon)$ independent of η_i and v_{is} for all t, s and kept the observations on x_{it} fixed over replications. As an alternative choice of regressors we used total sales from a sample of quoted U.K. firms where large variations across units and outliers are likely to be present. In both cases, x_{it} is strictly exogenous and uncorrelated with the individual effects. However, since we are interested in the performance of estimators that rely on lags of y_{it} for the identification of α and β, the over-identifying restrictions arising from the strict exogeneity of x_{it} and the lack of correlation with η_i were not used in the simulated GMM estimator. Thus, we chose[6]

$$Z_i = [\mathrm{diag}\,(y_{i1} \cdots y_{is}) \vdots (\bar{x}_{i3} \cdots \bar{x}_{iT})'] \qquad (s = 1, \ldots, T-2)$$

which is a valid instrument set provided $\phi = 0$.

In the base design, the sample size is $N = 100$ and $T = 7$, the v_{it} are independent over time and homoskedastic: $\theta_1 = \phi = 0$, $\theta_0 = \sigma^2 = 1$, $\sigma^2_\eta = 1$, $\beta = 1$, $\rho = 0{\cdot}8$ and $\sigma^2_\varepsilon = 0{\cdot}9$. Tables 1 and 2 summarize the results for $\alpha = 0{\cdot}2$, $0{\cdot}5$, $0{\cdot}8$ obtained from 100 replications. Results for other variants of this design were calculated ($N = 200$, $T = 6, \sigma^2 = 2$, 5, $\sigma^2_\eta = 0$, $\rho = 0$], and are available from the authors on request. However the conclusions are the same as for the results reported here.

Table 1 reports sample means and standard deviations for one-step and two-step GMM estimators (GMM1 and GMM2 respectively), OLS in levels, within-groups, and

TABLE 1

Biases in the estimates

	GMM1	GMM2	OLS	Within-groups	AHd	AHl	One-step ASE	Robust One-step ASE	Two-step ASE
				$\alpha = 0{\cdot}5, \beta = 1$					
Coefficient: α									
Mean	0·4884	0·4920	0·7216	0·3954	−2·4753	0·5075	0·0683	0·0677	0·0604
St. Dev.	0·0671	0·0739	0·0216	0·0272	45·9859	0·0821	0·0096	0·0120	0·0106
Coefficient: β									
Mean	1·0053	0·9976	0·7002	1·0409	0·1625	0·9996	0·0612	0·0607	0·0548
St. Dev.	0·0631	0·0668	0·0484	0·0480	9·8406	0·0650	0·0031	0·0055	0·0052
				$\alpha = 0{\cdot}2, \beta = 1$					
Coefficient: α									
Mean	0·1937	0·1979	0·5108	0·0957	0·2025	0·2044	0·0610	0·0602	0·0533
St. Dev.	0·0597	0·0670	0·0340	0·0309	0·1973	0·0661	0·0045	0·0066	0·0060
Coefficient: β									
Mean	1·0048	0·9960	0·7030	1·0430	0·9973	0·9991	0·0620	0·0615	0·0553
St. Dev.	0·0630	0·0687	0·0526	0·0476	0·0818	0·0654	0·0028	0·0058	0·0052
				$\alpha = 0{\cdot}8, \beta = 1$					
Coefficient: α									
Mean	0·7827	0·7810	0·8997	0·7160	0·8103	0·8038	0·0529	0·0527	0·0470
St. Dev.	0·0582	0·0609	0·0090	0·0206	0·1313	0·2677	0·0069	0·0082	0·0075
Coefficient: β									
Mean	1·0001	0·9926	0·7754	1·0137	1·0000	0·9980	0·0609	0·0601	0·0544
St. Dev.	0·0622	0·0671	0·0423	0·0461	0·0789	0·0893	0·0035	0·0056	0·0056

Notes.

(i) $N = 100$, $T = 7$, 100 replications, $\sigma^2 = \sigma^2_\eta = 1$.

(ii) Exogenous variable is first order autoregressive with $\rho = 0{\cdot}8$ and $\sigma^2_\varepsilon = 0{\cdot}9$.

(iii) GMM1 and GMM2 are respectively one step and two step difference—IV estimators of the type described in Section 2. Both GMM use $Z_i = [\mathrm{diag}\,(y_{i1} \cdots y_{is}) : (\bar{x}_{i3} \cdots \bar{x}_{iT})']$ $(s = 1, \ldots, T-2)$.

(iv) AHd and AHl are the Anderson-Hsiao stacked—IV estimators of the equation in first differences that use $\Delta y_{i(t-2)}$ and $y_{i(t-2)}$ as an instrument for $\Delta y_{i(t-1)}$ respectively.

(v) One Step ASE and Robust One Step ASE are estimates of the asymptotic standard errors of GMM1. The former are only valid for i.i.d. errors while the latter are robust to general heteroskedasticity over individuals and over time. Two step ASE is a robust estimate of the asymptotic standard errors of GMM2.

6. The optimal instrument set for the system of first difference equations would be $z_i = \mathrm{diag}\,(y_{i1} \cdots y_{is}, x_{i1}, \ldots, x_{iT})$.

two AH estimators. The AHd estimator is given by

$$\begin{pmatrix}\hat{\alpha}\\ \hat{\beta}\end{pmatrix}_{\text{AHd}} = \left(\sum_i^N \sum_{t=4}^T \Delta z_{it} \Delta w'_{it}\right)^{-1} \sum_i^N \sum_{t=4}^T \Delta z_{it}\, \Delta y_{it} \tag{15}$$

where $w_{it} = (y_{i(t-1)}, x_{it})'$ and $z_{it} = (y_{i(t-2)}, x_{it})'$. The AHl estimator replaces Δz_{it} with $(y_{i(t-2)}, \Delta x_{it})'$ and the summation goes from $t=3$ to T. The next two columns report sample means and standard deviations for two alternative estimates of the asymptotic standard errors of the one step GMM estimator. The first one is only valid for i.i.d. errors while the second is robust to heteroskedasticity of arbitrary form. The last column corresponds to estimates of the asymptotic standard errors of the two-step GMM estimator.

The tabulated results show a small downward finite-sample bias in the GMM estimators of α (of about 2 to 3%). Not surprisingly, the OLS and the within-group (WG) estimators of α exhibit large biases in opposite directions (i.e. upward bias in OLS whose size depends on σ_η^2; downward bias in WG whose size depends on T). The behaviour of the AH estimators is more surprising. Concerning AHd, there is evidence of lack of identification for $\alpha = 0{\cdot}5$ and negligible biases, though coupled with large variances, for $\alpha = 0{\cdot}2$ and $0{\cdot}8$. On the other hand, the standard deviation of AHl is small for $\alpha = 0{\cdot}2$ and $\alpha = 0{\cdot}5$ but it more than doubles that of AHd for $\alpha = 0{\cdot}8$. These results are consistent with the calculations of asymptotic variance matrices for the AH estimators reported elsewhere (cf. Arellano (1989)). As explained in that note, in a model containing an exogenous variable in addition to the lagged dependent variable, there are values of α and ρ between 0 and 1 for which there is no correlation between $\Delta y_{i(t-1)}$ and $\Delta y_{i(t-2)}$, in which case $\Delta y_{i(t-2)}$ is not a valid instrument and AHd is not identified. In our first experiment, AHd is close to such a singularity which explains the result. In contrast, AHl has no singularities for stationary values of α and ρ but can nevertheless be even less precise than AHd for large values of α.

An interesting result is that the standard deviation of the GMM estimators of α is about three times smaller than that of AHd for $\alpha = 0{\cdot}2$ and $0{\cdot}8$ and between four and five times smaller than that of AHl for $\alpha = 0{\cdot}8$ (although the standard deviation of AHl for $\alpha = 0{\cdot}2$ and $\alpha = 0{\cdot}5$ is of a similar magnitude as for the GMM estimators). This suggests that there may be significant efficiency gains in practice by using GMM as opposed to AH, aside from overcoming potential singularities as in our first experiment.

Concerning GMM1, the two alternative estimators of their asymptotic standard errors behave in a similar way, although the robust alternatives always have a bigger standard deviation. Their sample mean is always very close to the finite-sample standard deviation in column one, suggesting that the asymptotic approximation is quite accurate for the simulated designs. On the other hand, the estimator of the asymptotic standard errors of GMM2 in the last column shows a downward bias of around 20 percent relative to the finite-sample standard deviations reported in the second column.

Table 2 reports the number of rejections together with sample means and variances for the test statistics discussed in Section 3. The first three columns contain two alternative versions of the one step m_2 statistic and the two step m_2 statistic (see the notes to the table). The Sargan tests are tests of the over-identifying restrictions based on minimized criteria of the GMM estimators of Table 1. The difference-Sargan tests are based on the difference between the minimized GMM criteria and the restricted versions of these that remain valid when the errors are MA (1). The Hausman statistics test the distance between the GMM and the restricted GMM estimates of α.

With only 100 replications we cannot hope to provide accurate estimates of the tail probabilities associated with the test statistics; our results can only be suggestive. The

TABLE 2

Sizes of the Test Statistics, Number of rejections out of 100 cases

	One-step m_2	Robust one-step m_2	Two-step m_2	One-step Sargan (df = 14)	Two-step Sargan (df = 14)	One-step difference-Sargan (df = 5)	Two-step difference-Sargan (df = 5)	One-step Hausman (df = 1)	Two-step Hausman (df = 1)
					$\alpha = 0{\cdot}5$				
10	12	15	14	11	7	13	12	13	20
5	5	5	6	6	4	6	6	9	12
1	1	1	1	0	0	0	2	2	3
Mean	0·013	0·003	0·002	13·622	13·844	5·195	5·209	1·135	1·487
Variance	1·049	1·071	1·063	28·004	23·092	10·889	10·788	3·547	4·693
					$\alpha = 0{\cdot}2$				
10	12	13	13	8	6	12	9	11	17
5	4	5	4	4	3	8	6	7	13
1	2	1	1	0	0	1	1	2	3
Mean	−0·002	−0·010	−0·018	13·519	13·691	5·043	5·004	1·047	1·528
Variance	1·127	1·146	1·124	24·471	22·092	10·928	10·806	2·612	5·003
					$\alpha = 0{\cdot}8$				
10	12	15	15	9	10	7	12	11	12
5	5	6	6	7	4	4	9	5	8
1	1	0	0	1	2	1	2	0	4
Mean	0·047	0·037	0·036	13·423	13·883	4·917	5·199	0·928	1·239
Variance	1·004	1·029	1·028	31·492	28·854	8·916	11·333	1·648	4·325

Notes.

(i) $N = 100$, $T = 7$, 100 replications, $\sigma^2 = \sigma_\eta^2 = 1$, $\beta = 1$.

(ii) Exogenous variable is first order autoregressive with $\rho = 0{\cdot}8$ and $\sigma_\varepsilon^2 = 0{\cdot}9$.

(iii) All tests are based on the one- and two-step GMM estimates reported in Table 1 as well as on restricted versions of those which only use the columns of Z_i that remain valid instruments when the errors are MA (1). The one-step m_2 statistic is described in the Appendix.

(iv) Results on an extended set of experiments are available from the authors on request.

TABLE 3

Power of the Test Statistics, Number of rejections out of 100 cases

	One-step m_2	Robust one-step m_2	Two-step m_2	One-step Sargan (df = 14)	Two-step Sargan (df = 14)	One-step difference-Sargan (df = 5)	Two-step difference-Sargan (df = 5)	One-step Hausman (df = 1)	Two-step Hausman (df = 1)
Serial correlation: Corr $(v_t v_{t-1}) = 0 \cdot 2$									
10	54	53	53	47	33	60	46	27	25
5	45	46	46	33	20	53	34	15	20
1	24	25	25	22	2	29	9	8	11
Mean	−1·833	−1·823	−1·823	20·673	17·988	11·307	9·151	2·076	2·431
Variance	1·079	0·987	0·988	70·881	34·447	40·017	20·773	8·855	12·496
Serial correlation: Corr $(v_t v_{t-1}) = 0 \cdot 3$									
10	95	96	96	78	71	91	86	29	33
5	91	92	92	72	47	86	70	20	22
1	78	77	78	57	30	69	51	11	11
Mean	−3·293	−3·118	−3·121	32·057	24·180	22·112	15·314	2·537	2·707
Variance	1·033	0·695	0·705	135·543	40·127	94·261	30·339	12·430	13·640
Heteroskedasticity: Var $(v_{it}) = x_{it}^2$, x_{it} AR (1) data									
10	23	9	10	43	8	35	7	24	25
5	15	2	2	35	0	24	4	16	16
1	3	0	0	13	0	15	0	8	5
Mean	0·029	−0·023	−0·038	20·967	14·495	9·104	5·216	2·158	1·938
Variance	1·708	1·005	0·991	70·077	17·194	36·410	8·737	13·954	6·355
Heteroskedasticity: Var $(v_{it}) = x_{it}^2$, x_{it} : U.K. Sales									
10	71	6	7	100	0	93	13	74	81
5	68	3	2	100	0	92	3	69	76
1	54	0	0	100	0	88	1	62	72
Mean	−0·870	−0·225	−0·301	207·789	14·940	63·389	5·611	43·241	154·452
Variance	15·022	1·126	1·014	5863·623	5·405	2659·712	9·095	12059·642	339856·703

Notes.

(i) $N = 100$, $T = 7$, 100 replications, $\alpha = 0 \cdot 5$, $\beta = 1$, $\sigma_\eta^2 = 0$.

(ii) In the serial correlation designs v_{it} has been generated as MA (1) with standard normal random errors.

(iii) In the first three designs x_{it} is AR (1) with $\rho = 0 \cdot 8$ and $\sigma_\varepsilon^2 = 0 \cdot 9$.

(iv) In the fourth design, x_{it} are total sales from a sample of quoted U.K. companies scaled to have Var $(\Delta x_{it}) = 1$. x_{it} behaves as a random walk process with large differences in mean and variance across firms.

(v) In the heteroskedastic designs there is no serial correlation in the v_{it}.

robust m_2 statistics, which depend on the fourth-order moments of the data, both tend to reject too often at the 10% level, suggesting that they have a slower convergence to normality by comparison with the other test, but they are still to be recommended when heteroskedasticity is suspected. Overall all three m_2 tests seem to be quite well approximated by their asymptotic distributions under the null, with no obvious indications of the need for systematic finite-sample size corrections. The same is true for the Sargan, difference-Sargan and one step Hausman tests. However, the two-step Hausman statistic appears to over-reject consistently in these experiments.

Table 3 repeats the exercise for two models with MA (1) serial correlation ($\phi = 0{\cdot}209$ and $0{\cdot}333$) and two other experiments with heteroskedastic errors. The m_2 statistic will reject the null more than half the time at the 10 per cent level when the correlation between v_{it} and $v_{i(t-1)}$ is only 0·2. However when the autocorrelation rises to 0·3, the null will be rejected in 95% of cases. The Hausman test has considerably less power than the difference-Sargan test or the m_2 statistics, and with increasing autocorrelation the difference in power becomes wider.

The last two panels of Table 3 investigate the effects of departures from homoskedasticity of the error distribution on the probabilities of rejection of the tests. Both experiments have $\theta_0 = 0$ and $\theta_1 = 1$. In the first, the x_{it} are generated AR (1) data as in the previous experiments, while in the second the x_{it} are U.K. sales data. This has a dramatic effect on the one-step tests which are not robust to heteroskedasticity. On the other hand, the robust m_2 statistics and the two-step difference-Sargan test show no serious departures from their nominal size. The two-step Sargan test tends to under-reject and the two-step Hausman test over-rejects, especially in the last experiment where the variance of x_{it} is much greater. We suspect that the two-step Hausman statistic is very sensitive to the presence of outliers.

5. AN APPLICATION TO EMPLOYMENT EQUATIONS

In this section we apply the strategy for estimation and testing outlined earlier to a model of employment, using panel data for a sample of U.K. companies. We consider a dynamic employment equation of the form

$$n_{it} = \alpha_1 n_{i(t-1)} + \alpha_2 n_{i(t-2)} + \beta'(L) x_{it} + \lambda_t + \eta_i + v_{it}. \tag{16}$$

Here n_{it} is the logarithm of U.K. employment in company i at the end of year t,[7] the vector x_{it} contains a set of explanatory variables and $\beta(L)$ is a vector of polynomials in the lag operator. The specification also contains a time effect λ_t that is common to all companies,[8] a permanent but unobservable firm-specific effect η_i and an error term v_{it}.

Equation (16) will admit more than one theoretical interpretation. Suppose first that, in the absence of adjustment costs, a price-setting firm facing a constant elasticity demand curve would choose to set employment according to a log-linear labour demand equation (see, for example, Layard and Nickell (1986))

$$n^*_{it} = \gamma_0 + \gamma_1 w_{it} + \gamma_2 k_{it} + \gamma_3 \sigma^e_{it} + \eta'_i \tag{17}$$

where $\gamma_1 < 0$, $\gamma_2 > 0$ and $\gamma_3 \geqq 0$. Here w_{it} is the log of the real product wage, k_{it} is the log

7. Note that the time period is taken to be the 12-month period covered in the company's accounts ("accounting year") and so differs across companies in the sample.

8. These time effects relate to calendar years and a company's accounting year is allocated to the calendar year in which it ends.

of gross capital, σ^e_{it} is a measure of expected demand for the firm's product relative to potential output, and the intercept may contain a firm-specific component η'_i. If employment adjustment is costly then actual employment will deviate from n^*_{it} in the short run. This suggests a dynamic labour demand model of the form of (16), where x_{it} contains k_{it}, w_{it} and σ^2_{it}, and unrestricted lag structures are included to model this sluggish adjustment. We include the log of industry output (ys_{it}) to capture industry demand shocks, and aggregate demand shocks are also included through the time dummies. The resulting employment equation is a skeleton version of those estimated on U.K. time series data by Layard and Nickell (1986) and on micro data by Nickell and Wadhwani (1989). The short-run dynamics will compound influences from adjustment costs, expectations formation and decision processes.

Alternatively, if adjustment costs take the standard additively-separable quadratic form $(1/2a)(N_{it}-N_{i(t-1)})^2$, where N_{it} denotes the level rather than the logarithm of company employment, then Dolado (1987), following Nickell (1984), derives a log-linear approximation to the Euler equation for a firm maximising the present discounted value of profits as

$$E_{t-1}(n_{it}) = \delta_0 + (2+r)n_{i(t-1)} - (1+r)n_{i(t-2)} + a(1+r)[n_{i(t-1)} - n^*_{i(t-1)}]. \tag{18}$$

Here r is a real discount rate, assumed constant, and n^*_{it} is given by (17). Replacing the conditional expectation by its realisation and introducing an expectational error v_{it} yields a model with the form of (16), though with strong restrictions on the dynamic structure in this case. In particular the rational expectations hypothesis suggests a theoretical motivation for the assumption of serially-uncorrelated errors in this kind of model.

The principal data source used is the published accounts of 140 quoted companies whose main activity is manufacturing and for which we have seven or more continuous observations during the period 1976–1984. The panel is unbalanced both in the sense that we have more observations on some firms than on others, and because these observations correspond to different points in historical time. We allocate each of our companies to one of nine broad sub-sectors of manufacturing according to their main product by sales, and use value-added in that sector as our measure of industry output. Our wage variable is a measure of average remuneration per employee in the company, which we deflate using a value-added price deflator at the industry level. Finally we use an inflation-adjusted estimate of the company's gross capital stock. More information about the sample and the construction of these variables is provided in the Data Appendix.

In Table 4 we report GMM estimates of these dynamic employment equations.[9] We begin by including current-dated variables and unrestricted lag structures. Columns (a1) and (a2) present the one-step and two-step results respectively for the most general dynamic specification that we considered. Three cross-sections are lost in constructing lags and taking first differences, so that the estimation period is 1979–1984, with 611 useable observations. Here all variables other than the lagged dependent variables are assumed to be strictly exogenous, although none of the over-identifying restrictions that follow from this assumption are exploited.

Comparing columns (a1) and (a2) shows that the estimated coefficients are quite similar in all cases. Both models are well determined and have sensible long-run properties for a labour demand equation. However the asymptotic standard errors associated with the two-step estimates are generally around 30% lower than those associated with the

9. Estimation was performed using the DPD program written in GAUSS, described in Arellano and Bond (1988a) and available from the authors on request.

TABLE 4

Employment equations

GMM estimates (all variables in first differences)

Dependent variable: ln (Employment)$_{it}$ Sample Period: 1979–1984 (140 companies)

Independent variables	(a1)	(a2)	(b)	Instrumenting wages and capital* (c)	(d)
$n_{i(t-1)}$	0·686 (0·145)	0·629 (0·090)	0·474 (0·085)	0·800 (0.048)	0·825 (0.056)
$n_{i(t-2)}$	−0·085 (0·056)	−0·065 (0·027)	−0·053 (0·027)	−0·116 (0·021)	−0·074 (0·020)
w_{it}	−0·608 (0·178)	−0·526 (0·054)	−0·513 (0·049)	−0·640 (0·054)	—
$w_{i(t-1)}$	0·393 (0·168)	0·311 (0·094)	0·225 (0·080)	0·564 (0·066)	0·431 (0·076)
k_{it}	0·357 (0·059)	0·278 (0·045)	0·293 (0·039)	0·220 (0·051)	—
$k_{i(t-1)}$	−0·058 (0·073)	0·014 (0·053)	—	—	−0·077 (0·045)
$k_{i(t-2)}$	−0·020 (0·033)	−0·040 (0·026)	—	—	—
ys_{it}	0·608 (0.172)	0·592 (0·116)	0·610 (0·109)	0·890 (0·098)	—
$ys_{i(t-1)}$	−0·711 (0·232)	−0·566 (0·140)	−0·446 (0·125)	−0·875 (0·105)	−0·115(0·100)
$ys_{i(t-2)}$	0·106 (0·141)	0·101 (0·113)	—	—	0·096 (0·092)
m_2	−0·516	−0·434	−0·327	−0·610	−1·259
Sargan test	65·8 (25)	31·4 (25)	30·1 (25)	63·0 (50)	68·3 (51)
Difference-Sargan	41·9 (6)	15·4 (6)	10·0 (6)	28·6 (20)	31·6 (20)
Hausman	5·8 (1)	14·4 (1)	13·4 (1)	2·0 (1)	2·9 (1)
Wald test	408·3 (10)	667·0 (10)	372·0 (7)	779·3 (7)	623·9 (6)
No. of observations	611	611	611	611	611

* A subset of valid moment restrictions involving lagged wages and capital are exploited—see note (vi). Additional instruments used are the stacked levels and first differences of (firm real sales)$_{(t-2)}$ and (firm real stocks)$_{(t-2)}$.

Notes

(i) Time dummies are included in all equations.

(ii) Asymptotic standard errors robust to general cross-section and time series heteroskedasticity are reported in parentheses.

(iii) The GMM estimates reported are all two step except column (a1).

(iv) The m_2, Sargan, difference-Sargan and Hausman statistics are all two step versions of these tests except in column (a1). In column (a1) the m_2 and Hausman statistics are asymptotically robust to general heteroskedasticity, whilst the Sargan and difference-Sargan tests are only valid in the case of i.i.d. errors. All Hausman statistics test only the coefficient on $n_{i(t-1)}$. Degrees of freedom for χ^2 statistics are reported in parentheses.

(v) The Wald statistic is a test of the joint significance of the independent variables asymptotically distributed as χ^2_k under the null of no relationship, where k is the number of coefficients estimated (excluding time dummies).

(vi) The basic instrument set used in columns (a1), (a2) and (b) is of the form

$$Z_i \begin{bmatrix} n_{i1} & n_{i2} & 0 & 0 & 0 & \cdots & 0 & \cdots & 0 & \vdots & \Delta x'_{i4} \\ 0 & 0 & n_{i1} & n_{i2} & n_{i3} & & 0 & & 0 & \vdots & \Delta x'_{i5} \\ \vdots & & & & & \ddots & & & & & \\ 0 & 0 & 0 & 0 & 0 & \cdots & n_{i1} & \cdots & n_{i7} & . & \Delta x'_{i9} \end{bmatrix} \begin{matrix} 1979 \\ 1980 \\ \\ 1984 \end{matrix}$$

where x_{it} is the vector of exogenous variables included in the equation. For example, the equation for 1979 in first differences can be written as

$$\Delta n_{i4} = \alpha_1 \Delta n_{i3} + \alpha_2 \Delta n_{i2} + \Delta x'_{i4}\beta + \Delta v_{i4}$$

For companies on which less than 9 observations are available, the rows of Z_i corresponding to the missing equations are deleted and the missing values of n in the remaining rows are replaced by zeroes.

In columns (c) and (d) Z_i is modified to take the form

$$Z_i = [\text{diag}\,(n_{i1} \cdots n_{is} w_{i(s-1)} w_{is} k_{i(s-1)} k_{is}) \vdots (\Delta x'_{i4} \cdots \Delta x'_{i9})'] \qquad (s = 2, \ldots, 7)$$

where x_{it} is now the vector of explanatory variables excluding wages and capital but including stacked lagged sales and stocks.

one-step estimates, with the discrepancy being even larger in some cases. We suspect that most of this apparent gain in precision may reflect a downward finite-sample bias in the estimates of the two-step standard errors as indicated by the simulation results in Table 1, suggesting that caution would be advisable in making inferences based on the two-step estimator alone in samples of this size.

Turning to the test statistics, neither of the robust m_2 statistics nor the two-step Sargan test provide evidence to suggest that the assumption of serially uncorrelated errors is inappropriate in this example. The one-step Sargan and difference-Sargan statistics do reject the overidentifying restrictions but our simulation results showed a strong tendency for those tests to reject too often in the presence of heteroskedasticity. The two-step difference-Sargan test is more marginal but does reject at the 5 per cent significance level. Both Hausman tests also reject but these too show a tendency to over-reject in our simulation experiments. One possibility is that this instability across different instrument sets reflects the failure of the strict exogeneity assumption for wages and capital, rather than serial correlation per se.

In Table 5 we present some alternative estimates of this same model for comparison. Columns (e) and (f) report two instrumental variable estimates of the differenced equation using simpler instrument sets of the AH type. In column (e) we use the difference $\Delta n_{i(t-3)}$ to instrument $\Delta n_{i(t-1)}$, losing one further cross-section, whilst in column (f) we use the level $n_{i(t-3)}$ as the instrument. In both cases the coefficient estimates are poorly determined, indicating a massive loss in efficiency compared to either GMM estimator in this application. Using both $\Delta n_{i(t-3)}$ and $n_{i(t-3)}$ as instruments (not reported) helped a little, but the estimates remained very imprecise. In column (g) we report OLS estimates of the employment equation in levels. In this case the 1978 cross-section is available and the longer estimation period has been used here. Compared to the GMM estimates there is a serious upward bias on the lagged dependent variable, which suggests the presence of firm-specific effects. Column (h) reports the within-groups estimates, which are close to GMM in this example. In fact the WG estimate of the first-order autoregressive coefficient is bigger than the corresponding GMM estimates, although the comparison between WG and GMM in this case is obscured by the likely endogeneity of wages and capital.

Returning to the GMM estimates in Table 4, column (b) omits insignificant dynamics with little change in the long-run properties of the previous model. In columns (b)–(d) we report only the two-step estimates though the one-step coefficient estimates were invariably similar. In column (b) the two-step difference-Sargan test now marginally accepts the hypothesis of no serial correlation, but the two-step Hausman statistic remains an outlier. In column (c) we relax the assumption that the real wage and capital stock are strictly exogenous and instead treat them as being endogenous. We therefore use lags of w and k dated $(t-2)$ and earlier as instruments for w_{it}, $w_{i(t-1)}$ and k_{it}. We also use lagged values of the company's real sales and real stocks as additional instruments. Given the size of our sample, not all the available moment restrictions were used. The precise form of the instrument matrix is described in note (vi) to Table 4. The results in column (c) suggest that it is inappropriate to treat wages and capital as strictly exogenous in this model. In this case none of the test statistics indicate the presence of misspecification.

The coefficient estimates for our preferred specification in column (c) suggest a long-run wage elasticity of $-0{\cdot}24$ (standard error $= 0{\cdot}28$) and a long-run elasticity with respect to capital of $0{\cdot}7$ (S.E. $= 0{\cdot}14$). There is a strong suggestion that industry output enters the model in changes rather than levels, which is appealing since σ^e_{it} in (17) measures demand shocks relative to potential output. Layard and Nickell (1986) interpret

TABLE 5

Employment equations

Alternative estimates

Dependent variable: ln (Employment)$_{it}$

Independent variables	(e) AHd	(f) AHl	(g) OLS	(h) Within-groups
$n_{i(t-1)}$	1·423	2·308	1·045	0·734
	(1·001)	(1·055)	(0·051)	(0·058)
$n_{i(t-2)}$	−0·165	−0·224	−0·077	−0·141
	(0·128)	(0·117)	(0·048)	(0·077)
w_{it}	−0·752	−0·810	−0·524	−0·557
	(0·230)	(0·283)	(0·172)	(0·155)
$w_{i(t-1)}$	0·963	1·422	0·477	0·326
	(0·768)	(0·851)	(0·169)	(0·143)
k_{it}	0·322	0·253	0·343	0·385
	(0·105)	(0·110)	(0·048)	(0·056)
$k_{i(t-1)}$	−0·325	−0·552	−0·202	−0·084
	(0·386)	(0·357)	(0·064)	(0·053)
$k_{i(t-2)}$	−0·095	−0·213	−0·116	−0·025
	(0·123)	(0·145)	(0·035)	(0·042)
ys_{it}	0·766	0·991	0·433	0·521
	(0·311)	(0·338)	(0·176)	(0·193)
$ys_{i(t-1)}$	−1·362	−1·938	−0·768	−0·659
	(0·881)	(0·992)	(0·248)	(0·208)
$ys_{i(t-2)}$	0·321	0·487	0·312	0·001
	(0·416)	(0·425)	(0·130)	(0·139)
m_2	−0·781	−0·919	−1·029	
Wald test	199·3 (10)	101·1 (10)		
R^2			0·994	0·689
Number of observations	471	611	751	751

Notes.

(i) Time dummies are included in all equations.

(ii) Asymptotic standard errors robust to general cross-section and time series heteroskedasticity are reported in parentheses.

(iii) The m_2 and Wald tests are asymptotically robust to general heteroskedasticity.

(iv) Columns (e) and (f) report Anderson-Hsiao-type estimates of the equation in first differences: $\Delta n_{i(t-1)}$ is treated as endogenous and the additional instruments used are $\Delta n_{i(t-3)}$ in (e), and $n_{i(t-3)}$ in (f), so that one further cross-section is lost in (e) and the effective sample period becomes 1980–84.

(v) Column (g) reports OLS estimates of the equation in levels, where the effective sample period becomes 1978–1984.

(vi) Column (h) reports within-groups estimates. These are OLS estimates of the equation in deviations from time means.

the short-run effect of product demand fluctuations on labour demand as reflecting the practice of normal cost pricing.

Finally in column (d) we report estimates of the Euler equation model given in (18). Here again we treat wages and capital as endogenous variables. Although the tests for serial correlation remain below their critical values, the coefficient estimates are not favourable to the Euler equation interpretation. The coefficients on capital and industry output are poorly determined, whilst those on the lagged dependent variable imply a real discount rate of around −100%. Very similar results were obtained for versions of the Euler equation model allowing for an MA (1) error process and estimating in levels as opposed to logs. It appears that the process of employment adjustment is not well described by this model.

The results of this empirical application are generally in agreement with those of our Monte Carlo simulations. The GMM estimator offers significant efficiency gains compared to simpler IV alternatives, and produces estimates that are well-determined in dynamic panel data models. The tendency for non-robust test statistics to over-reject is confirmed. The robust m_2 statistics perform satisfactorily as do the two-step Sargan and difference-Sargan tests, but the two-step Hausman test must be considered suspect in samples of this size.

6. CONCLUSION

In this paper we have discussed the estimation of dynamic panel data models by the generalized method of moments. The estimators we consider exploit optimally all the linear moment restrictions that follow from particular specifications, and are extended to cover the case of unbalanced panel data. We focus on models with predetermined but not strictly exogenous explanatory variables in which identification results from lack of serial correlation in the errors. A test of serial correlation based on the GMM residuals is proposed and compared with Sargan tests of over-identifying restrictions and Hausman specification tests.

To study the practical performance of these procedures we performed a Monte Carlo simulation for 100 units, seven time-periods and two parameters. The results indicate negligible finite sample biases in the GMM estimators and substantially smaller variances than those associated with simpler IV estimators of the kind introduced by Anderson and Hsiao (1981). We also find that the distributions of the serial-correlation tests are well-approximated by their asymptotic counterparts.

We applied these methods to estimate employment equations using an unbalanced panel of 140 quoted U.K. companies for the period 1979–1984. The GMM estimators and the serial correlation tests performed well in this application. A potentially serious problem, suggested by both the experimental evidence and the application, concerns the estimates of the standard errors for the two-step GMM estimator which we find to be downward biased in our samples. Further results on alternative estimators of these standard errors would be very useful.

APPENDIX

A. The asymptotic normality of the m_2 statistic

Following the notation of Section 3, under the assumption that $(X'_{-2}v_*/N) = o_p(1)$ we have

$$N^{-1/2}\hat{v}'_{-2}\hat{v}_* = N^{-1/2}v'_{-2}v_* - (v'_{-2}X_*/N)N^{1/2}(\hat{\delta}-\delta) + o_p(1) \quad \text{as } N \to \infty$$

and also

$$N^{-1/2}\hat{v}'_{-2}\hat{v}_* = N^{-1/2}v'_{-2}v_* - g'_N N^{-1/2}Z'v + o_p(1),$$

where

$$g'_N = v'_{-2}X_*(X'ZA_NZ'X)^{-1}X'ZA_N.$$

Then a multivariate central limit theorem for independent observations ensures

$$\bar{W}_N^{-1/2}N^{-1/2}\begin{pmatrix} v'_{-2}v_* \\ Z'v \end{pmatrix} \to^d N(0, I_{p+1}),$$

where $\bar{W}_N$ is the average covariance matrix of $(\phi_i, \xi_i')'$:

$$\bar{W}_N = \frac{1}{N}\sum_{i=1}^N E\begin{pmatrix} \phi_i^2 & \phi_i\xi_i' \\ \phi_i\xi_i & \xi_i\xi_i' \end{pmatrix} = \begin{pmatrix} \bar{\omega}_N & \bar{\psi}_N' \\ \bar{\psi}_N & \bar{V}_N \end{pmatrix},$$

and $\xi_i = Z_i' v_i$. Therefore

$$\bar{v}_N^{-1/2} N^{-1/2} \hat{v}_{-2}' v_* \to^d N(0, 1) \tag{A1}$$

where

$$\bar{v}_N = \bar{\omega}_N - 2g_N'\bar{\psi}_N + g_N'\bar{V}_N g_N$$

or

$$\bar{v}_N = \frac{1}{N}[\sum_{i=1}^N E(v_{i(-2)}' v_{i*} v_{i*}' v_{i(-2)}) - 2(v_{-2}'X_*)(X'ZA_NZ'X)^{-1}X'ZA_N\sum_{i=1}^N E(Z_i'v_iv_{i*}'v_{i(-2)})$$
$$+ (v_{-2}'X_*) \text{ avar } (\hat{\delta})(X_*'v_{-2})]$$

with

$$\text{avar } (\hat{\delta}) = N(X'ZA_NZ'X)^{-1}(X'ZA_N\bar{V}_NA_NZ'X)(X'ZA_NZ'X)^{-1}.$$

A consistent estimate of $\bar{v}_N$ can be obtained by replacing population average expectations of errors by sample averages of residuals. Finally, noticing that under our assumptions (A1) remains valid after this replacement, the result follows.

For a one-step $\hat{\delta}$, we can also consider an alternative m_2 criterion ("one step m_2") which relies on more restrictive auxiliary distributional assumptions. Assuming that the errors in the model in levels are independent and identically distributed across individuals and time, we have

$$\bar{\omega}_N = \frac{1}{N}\sigma^2\sum_{i=1}^N E(v_{i(-2)}' H_{i*} v_{i(-2)})$$

where H_{i*} is a $(T_i - 4)$ square matrix which has twos in the main diagonal, minus ones in the first subdiagonals and zeros otherwise. Moreover

$$\bar{\psi}_N = \frac{1}{N}\sigma^2\sum_{i=1}^N E(Z_i' H_{i\dagger} v_{i(-2)})$$

where $H_{i\dagger}$ is a $(T_i - 2) \times (T_i - 4)$ matrix with

$$H_{i\dagger} = \begin{pmatrix} H_{i**} \\ H_{i*} \end{pmatrix}$$

and H_{i**} is a $2 \times (T_i - 4)$ matrix with minus one in the (2, 1) position and zeroes elsewhere. Under such conditions $\hat{v}$ can be replaced by

$$\bar{v} = \hat{\sigma}^2\sum_{i=1}^N \hat{v}_{i(-2)}' H_{i*}\hat{v}_{-2}'X_*(X'ZA_NZ'X)^{-1}X'ZA_N\left(\sum_{i=1}^N Z_i'H_{i\dagger}\hat{v}_{i(-2)}\right)\hat{\sigma}^2$$
$$+ \hat{v}_{-2}'X_* \text{ av}\hat{\text{a}}\text{r } (\hat{\delta})X_*'\hat{v}_{-2}$$

where $\hat{\sigma}^2$ is an estimate of σ^2.

B. The Sargan difference test

Note that s in (10) can be re-written as

$$s = \frac{\hat{v}'Z^*}{\sqrt{N}}\left(\frac{1}{N}\sum_{i=1}^N Z_i^{*\prime}\hat{v}_i\hat{v}_i'Z_i^*\right)^{-1}\frac{Z^{*\prime}\hat{v}}{\sqrt{N}},$$

where $Z^* = ZH$ and H is a $p \times p$ linear transformation matrix such that

$$Z^* = (Z_I \mid Z_{II})$$

with

$$\frac{1}{N}\sum_{i=1}^N Z_{Ii}'\hat{v}_i\hat{v}_i'Z_{IIi} \to^p 0.$$

Letting

$$\left(\frac{1}{N}\sum_{i=1}^{N} Z_i^{*\prime}\hat{v}_i\hat{v}_i' Z_i^*\right)^{-1} = C_N C_N'$$

where C_N is $p \times p$ nonsingular, and noting that

$$\frac{Z^{*\prime}\hat{v}}{\sqrt{N}} = \left\{ I_p - \frac{Z^{*\prime}X}{N}\left[\frac{X'Z^*}{N} C_N C_N' \frac{Z^{*\prime}X}{N}\right]^{-1} \frac{X'Z^*}{N} C_N C_N' \right\} \frac{Z^{*\prime}v}{\sqrt{N}},$$

we have

$$s = \frac{v'Z^*}{\sqrt{N}} C_N[I_p - G(G'G)^{-1}G']C_N' \frac{Z^{*\prime}v}{\sqrt{N}}$$

where

$$G = C_N'\left(\frac{Z^{*\prime}X}{N}\right).$$

Following the usual argument one can show that $s \to^d \varepsilon' M \varepsilon \sim \chi^2_{p-k}$, where $\varepsilon \sim N(0, I_p)$ and M is of the form $I_p - D(D'D)^{-1}D'$ with rank $p-k$. O On the same lines we can write

$$s_I = \frac{v'Z_I}{\sqrt{N}} C_{IN}[I_{p_I} - G_I(G_I'G_I)^{-1}G_I']C_{IN}' \frac{Z_I'v}{\sqrt{N}},$$

where

$$G_I = C_{IN}'\left(\frac{Z_I'X}{N}\right)$$

and

$$\left(\frac{1}{N}\sum_{i=1}^{N} Z_{Ii}'\hat{v}_{Ii}\hat{v}_{Ii}' Z_{Ii}\right)^{-1} = C_{IN} C_{IN}'.$$

Let G^* contain the top p_I rows of G. Notice that $G^* - G_I \to^p 0$.
Therefore

$$ds = s - s_I \to^d \varepsilon' M \varepsilon - \varepsilon' \begin{pmatrix} M_I & 0 \\ 0 & 0 \end{pmatrix} \varepsilon$$

where $M_I = I_{p_I} - D_I(D_I'D_I)^{-1}D_I'$ and rank $(M_I) = p_I - k$ with $D' = (D_I' \mid D_{II}')$.
Finally notice that

$$\left[M - \begin{pmatrix} M_I & 0 \\ 0 & 0 \end{pmatrix}\right]$$

is symmetric and idempotent with rank $p - p_I$ and also

$$\left[M - \begin{pmatrix} M_I & 0 \\ 0 & 0 \end{pmatrix}\right]\begin{pmatrix} M_I & 0 \\ 0 & 0 \end{pmatrix} = 0,$$

from which (11) follows.

DATA APPENDIX

(a) Sample

The principal data source used is company accounts from Datastream International which provide accounts records of employment and remuneration (i.e. wage bill) for all U.K. quoted companies from 1976 onwards. We have used a sample of 140 companies with operations mainly in the U.K., whose main activity is manufacturing and for which we have at least 7 continuous observations during the period 1976-1984. Where more than 7 observations are available we have exploited this additional information, so that our sample has the unbalanced structure described in Table A1.

TABLE A1

Number of records on each company	Number of companies
7	103
8	23
9	14

As well as requiring at least 7 continuous observations, companies were excluded from our sample for a number of reasons. We required complete records on a set of accounting variables including gross fixed assets, investment, inventories and sales as well as employment and remuneration. Companies that changed the date of their accounting year end by more than a few days were excluded, so that our data all refer to 12 month periods. We also excluded companies where either employment or one of our constructed measures of real wages, real capital, real inventories or real sales jumped by more than a factor of 3 from one year to the next. This filter will remove both those companies where data has been recorded erroneously and those companies that have experienced major mergers. Finally we restricted our attention to companies that we could allocate to one of 9 broad sub-sectors of manufacturing industry using Datastream's breakdown of total sales by product available from 1980 onwards.

(*b*) *Variables*

Employment

Number of U.K. employees (Datastream variable 216)

Real Wage

A measure of average annual remuneration per employee was constructed by dividing U.K. remuneration (Datastream variable 214) by the number of U.K. employees. This was adjusted to take into account changes in average weekly hours worked in manufacturing industries (manual and non-manual employees, 18 years and over, male and female, all occupations—source: Department of Employment Gazette, various issues). A measure of real average hourly remuneration was then obtained by deflating using an implicit value-added price deflator. These implicit price deflators were calculated for each of our sub-sectors of manufacturing industry, using the current price and constant price GDP data published in various Blue Books.

Gross Capital Stock

Denoting the historic cost book value of gross fixed assets (Datastream variable 330) by HCK_t, we obtain an estimate of the inflation-adjusted (or replacement cost) value of gross fixed assets (RCK_t) using the formula

$$\text{RCK}_t = \text{HCK}_t \times \left(\frac{P_t^I}{P_{t-A}^I}\right),$$

where P^I is a price index for investment goods and A is an estimate of the average age of gross fixed assets. An implicit price deflator for gross fixed investment by manufacturing industry was calculated using the current price and constant price gross fixed investment data published in Economic Trends Annual Supplement (1986, p. 56). For the purpose of this exercise a value of A of 6 years was assumed. Our estimates of the gross capital stock at replacement cost are then expressed in constant prices using our investment goods deflator.

Industry Output

An index of value-added output at constant factor cost was constructed for each of our 9 sub-sectors of manufacturing industry, using data published in the Blue Book (1986, Table 2.4). The 15 sub-sectors of manufacturing for which this data is reported were combined into 9 using the weights given in the Blue Book.

Further details on this data set can be found in Arellano and Bond (1988b).

Acknowledgement. We are grateful to Charlie Bean, Richard Blundell, Olympia Bover, Michael Devereux, Zvi Griliches, David Hendry, Grayham Mizon, John Muellbauer, Steve Nickell, Alain Trognon, Sushil Wadhwani and two anonymous referees for helpful comments on this work. All remaining errors are our own. We thank the Institute for Fiscal Studies for their assistance and the ESRC for financial support under grant B00232207.

REFERENCES

AMEMIYA, T. and MACURDY, T. E. (1986), "Instrumental-Variable Estimation of an Error-Components Model", *Econometrica*, **54**, 869-881.

ANDERSON, T. W. and HSIAO, C. (1981), "Estimation of Dynamic Models with Error Components", *Journal of the American Statistical Association*, **76**, 598-606.

ANDERSON, T. W. and HSIAO, C. (1982), "Formulation and Estimation of Dynamic Models Using Panel Data", *Journal of Econometrics*, **18**, 47-82.

ARELLANO, M. (1989), "A Note on the Anderson-Hsiao Estimator for Panel Data", *Economics Letters*, **31**, 337-341.

ARELLANO, M. (1990), "Testing for Autocorrelation in Dynamic Random Effects Models", *Review of Economic Studies*, **57**, 127-134.

ARELLANO, M. and BOND, S. R. (1988a), "Dynamic Panel Data Estimation Using DPD—A Guide for Users" (Institute for Fiscal Studies, Working Paper 88/15, London).

ARELLANO, M. and BOND, S. R. (1988b), "Some Tests of Specification for Panel Data: Monte Carlo Evidence and an Application to Employment Equations" (Institute for Fiscal Studies, Working Paper 88/4, London).

BHARGAVA, A. and SARGAN, J. D. (1983), "Estimating Dynamic Random Effects Models from Panel Data Covering Short Time Periods", *Econometrica*, **51**, 1635-1659.

CHAMBERLAIN, G. (1984), "Panel Data", in Griliches, Z. and Intriligator, M. D. (eds.), *Handbook of Econometrics, Volume II.* (Amsterdam: Elsevier Science Publishers.)

DOLADO, J. J. (1987), "Intertemporal Employment and Pricing Decision Rules in UK Manufacturing" (University of Oxford, Applied Economics Discussion Paper 18).

GRILICHES, Z. AND HAUSMAN, J. A. (1986), "Errors in Variables in Panel Data", *Journal of Econometrics*, **31**, 93-118.

HANSEN, L. P. (1982), "Large Sample Properties of Generalized Method of Moments Estimators", *Econometrica*, **50**, 1029-1054.

HAUSMAN, J. A. (1978), "Specification Tests in Econometrics", *Econometrica*, **46**, 1251-1272.

HAUSMAN, J. A. and TAYLOR, W. E. (1981), "Panel Data and Unobservable Individual Effects", *Econometrica*, **49**, 1377-1398.

HOLTZ-EAKIN, D., NEWEY, W. and ROSEN, H. (1988), "Estimating Vector Autoregressions with Panel Data", *Econometrica*, **56**, 1371-1395.

HSIAO, C. (1986) *Analysis of Panel Data* (Cambridge: Cambridge University Press).

LAYARD, R. and NICKELL, S. J. (1986), "Unemployment in Britain", *Economica*, **53**, Supplement, 5121-5169.

NICKELL, S. J. (1984), "An Investigation of the Determinants of Manufacturing Employment in the United Kingdom", *Review of Economic Studies*, **51**, 529-557.

NICKELL, S. J. and WADHWANI, S. (1989), "Employment Determination in British Industry: Investigations Using Micro-Data" (Centre for Economic Policy Research Discussion Paper 320, London).

SARGAN, J. D. (1958), "The Estimation of Economic Relationships Using Instrumental Variables", *Econometrica*, **26**, 393-415.

SARGAN, J. D. (1980), "Some Tests of Dynamic Specification for a Single Equation", *Econometrica*, **48**, 879-897.

SARGAN, J. D. (1988), "Testing for Misspecification after Estimating Using Instrumental Variables", in Maasoumi, E. (ed.), *Contributions to Econometrics: John Denis Sargan, Vol.* 1 (Cambridge: Cambridge University Press).

WHITE, H. (1982), "Instrumental Variables Regression with Independent Observations", *Econometrica*, **50**, 483-499.