

GMM ESTIMATION WITH PERSISTENT PANEL DATA:
AN APPLICATION TO PRODUCTION FUNCTIONS

Richard Blundell

Stephen Bond

University College London
and Institute for Fiscal Studies

Nuffield College, Oxford
and Institute for Fiscal Studies

Key Words and Phrases: panel data; GMM; production functions

JEL classification: C23, D24

ABSTRACT

This paper considers the estimation of Cobb-Douglas production functions using panel data covering a large sample of companies observed for a small number of time periods. GMM estimators have been found to produce large finite-sample biases when using the standard first-differenced estimator. These biases can be dramatically reduced by exploiting reasonable stationarity restrictions on the initial conditions process. Using data for a panel of R&D-performing US manufacturing companies we find that the additional instruments used in our extended GMM estimator yield much more reasonable parameter estimates.

“In empirical practice, the application of panel methods to micro-data produced rather unsatisfactory results: low and often insignificant capital coefficients and unreasonably low estimates of returns to scale.”

— Griliches and Mairesse (1998).

1. Introduction

The estimation of simple Cobb-Douglas production functions from company panel data has become something of a graveyard for panel data estimation meth-

ods. As detailed in the recent paper by Griliches and Mairesse (1998), simple OLS regressions yield plausible parameter estimates, in line with evidence from factor shares and generally consistent with constant returns to scale. But attempts to control for unobserved heterogeneity and simultaneity - both likely sources of bias in the OLS results - have tended to yield less satisfactory parameter estimates. In particular, the application of GMM estimators which take first differences to eliminate unobserved firm-specific effects and use lagged instruments to correct for simultaneity in the first-differenced equations, has tended to produce very unsatisfactory results in this context (see, for example, Mairesse and Hall (1996)).

In this paper we suggest that these problems are related to the weak correlations that exist between the current growth rates of firm sales, capital and employment, and the lagged levels of these variables. This results in weak instruments in the context of the first-differenced GMM estimator. In an earlier paper (Blundell and Bond, 1998) we showed that weak instruments could cause large finite-sample biases when using the first-differenced GMM procedure to estimate autoregressive models for moderately persistent series from moderately short panels. We also showed that these biases could be dramatically reduced by incorporating more informative moment conditions that are valid under quite reasonable stationarity restrictions on the initial conditions process. Essentially this results in the use of lagged first-differences as instruments for equations in levels, *in addition to* the usual lagged levels as instruments for equations in first-differences (cf. Arellano and Bover, 1995).

Here we analyse whether similar issues are present in the production function application, and whether the extended GMM estimator gives more reasonable results in this context. Using a panel of R&D-performing US manufacturing firms similar to that used by Mairesse and Hall (1996), we first confirm that the first-differenced GMM estimator yields a low and statistically insignificant capital coefficient, and suggests sharply decreasing returns to scale. We then show that the sales, capital and employment series are highly persistent, and that the instruments used by the first-differenced estimator contain little information about the endogenous variables in first-differences. Using the extended GMM estimator, we find much more reasonable results: that is, we find a higher and strongly significant capital coefficient, and we do not reject constant returns to scale. The additional instruments used in this extended GMM estimator are not rejected in this application, and we confirm that the lagged first-differences are

informative instruments for the endogenous variables in levels. We also show that imposing constant returns to scale produces more reasonable results when the first-differenced GMM estimator is used. One further feature of our results is the importance of allowing for an AR(1) component in the production function error term. We need to allow for this serial correlation in order to obtain any valid lagged internal instruments for equations in first-differences or equations in levels.

The rest of the paper is organised as follows. Section 2 sets out the production function specification we estimate. Section 3 reviews the first-differenced GMM estimator, describes the extended ‘system’ GMM estimator, discusses the validity of the additional moment conditions which this estimator exploits in the production function context, and discusses how the validity of these additional moment conditions can be tested. Section 4 briefly describes the data we use, and Section 5 presents our empirical results. Section 6 concludes.

2. Model

We consider the Cobb-Douglas production function

$$y_{it} = \beta_n n_{it} + \beta_k k_{it} + \gamma_t + (\eta_i + v_{it} + m_{it}) \quad (2.1)$$

$$v_{it} = \rho v_{i,t-1} + e_{it} \quad |\rho| < 1$$

$$e_{it}, m_{it} \sim MA(0)$$

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.

We are interested in consistent estimation of the parameters (β_n, β_k, ρ) when the number of firms (N) is large and the number of years (T) is fixed. We maintain 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}).

The model has a dynamic (common factor) representation

$$y_{it} = \beta_n n_{it} - \rho \beta_n n_{i,t-1} + \beta_k k_{it} - \rho \beta_k k_{i,t-1} + \rho y_{i,t-1} + (\gamma_t - \rho \gamma_{t-1}) + (\eta_i(1 - \rho) + e_{it} + m_{it} - \rho m_{i,t-1}) \quad (2.2)$$

or

$$y_{it} = \pi_1 n_{it} + \pi_2 n_{i,t-1} + \pi_3 k_{it} + \pi_4 k_{i,t-1} + \pi_5 y_{i,t-1} + \gamma_t^* + (\eta_i^* + w_{it}) \quad (2.3)$$

subject to two non-linear (common factor) restrictions $\pi_2 = -\pi_1 \pi_5$ and $\pi_4 = -\pi_3 \pi_5$. Given consistent estimates of the unrestricted parameter vector $\pi = (\pi_1, \pi_2, \pi_3, \pi_4, \pi_5)$ and $\text{var}(\pi)$, these restrictions can be (tested and) imposed using minimum distance to obtain the restricted parameter vector (β_n, β_k, ρ) . Notice that $w_{it} = e_{it} \sim MA(0)$ if there are no measurement errors ($\text{var}(m_{it}) = 0$), and $w_{it} \sim MA(1)$ otherwise.

3. GMM estimation

3.1. First differences

A standard assumption on the initial conditions ($E[x_{i1}e_{it}] = E[x_{i1}m_{it}] = 0$ for $t = 2, \dots, T$) yields the following moment conditions

$$E[x_{i,t-s}\Delta w_{it}] = 0 \text{ where } x_{it} = (n_{it}, k_{it}, y_{it}) \quad (3.1)$$

for $s \geq 2$ when $w_{it} \sim MA(0)$, and for $s \geq 3$ when $w_{it} \sim MA(1)$. This allows the use of suitably lagged levels of the variables as instruments, after the equation has been first-differenced to eliminate the firm-specific effects (cf. Arellano and Bond, 1991).

Note however that the resulting first-differenced GMM estimator has been found to have poor finite sample properties (bias and imprecision) when the lagged levels of the series are only weakly correlated with subsequent first differences, so that the instruments available for the first-differenced equations are weak (cf. Blundell and Bond, 1998). This may arise here when the marginal processes for employment (n_{it}) and capital (k_{it}) are highly persistent, or close to random walk processes, as is often found to be the case.

To be more precise about these statements, consider the AR(1) model

$$y_{it} = \alpha y_{i,t-1} + \eta_i + v_{it} \quad |\alpha| < 1 \quad (3.2)$$

where v_{it} here is serially uncorrelated ($\rho = 0$). The instruments used in the standard first-differenced GMM estimator become less informative in two important cases. First, as the value of the autoregressive parameter α increases towards unity; and second, as the variance of the firm-specific effects (η_i) increases relative to the variance of the transitory shocks (v_{it}). For simplicity, consider the case with $T = 3$. In this case, the moment conditions corresponding to the first-differenced GMM estimator reduce to a single orthogonality condition. The first-differenced GMM estimator then corresponds to a simple instrumental variable (IV) estimator with reduced form (instrumental variable regression) equation

$$\Delta y_{i2} = \pi y_{i1} + r_i \text{ for } i = 1, \dots, N. \quad (3.3)$$

For sufficiently high autoregressive parameter α or for sufficiently high variance of the firm-specific effects, the least squares estimate of the reduced form coefficient π can be made arbitrarily close to zero (see Blundell and Bond, 1998). In this case the instrument y_{i1} is only weakly correlated with Δy_{i2} . We find that $\text{plim } \hat{\pi} \rightarrow 0$ as $\alpha \rightarrow 1$ or as $(\sigma_\eta^2/\sigma_v^2) \rightarrow \infty$, which are the cases in which the first stage F-statistic is $O_p(1)$.

Blundell and Bond (1998) characterise this problem of weak instruments using the concentration parameter of Nelson and Startz (1990a,b) and Staiger and Stock (1997). First note that the F-statistic for the first stage instrumental variable regression converges to a noncentral chi-squared with one degree of freedom. The concentration parameter is then the corresponding noncentrality parameter which we label τ . The IV estimator performs poorly when τ approaches zero. Assuming covariance stationarity, τ has the following simple characterisation in terms of the parameters of the AR(1) model

$$\tau = \frac{(\sigma_v^2 k)^2}{\sigma_\eta^2 + \sigma_v^2 k} \text{ where } k = \frac{(1 - \alpha)^2}{(1 - \alpha^2)}. \quad (3.4)$$

The performance of the first-differenced GMM estimator in this AR(1) specification can therefore be seen to deteriorate as $\alpha \rightarrow 1$, as well as for increasing values of $(\sigma_\eta^2/\sigma_v^2)$.

Blundell and Bond (1998) also report some results of a Monte Carlo study which investigates the finite sample properties of these GMM estimators in the AR(1) model. In Table I we present some specific examples that highlight the issues involved. We consider sample sizes with $N = 100$ and 500 , $T = 4$ and

TABLE I
Simulation Results

N	α	DIF	SYS
100	0.5	.4671 (.2674)	.5100 (.1330)
	0.8	.4844 (.8224)	.8101 (.1618)
	0.9	.2264 (.8264)	.9405 (.1564)
500	0.5	.4887 (.1172)	.5021 (.0632)
	0.8	.7386 (.3085)	.7939 (.0779)
	0.9	.5978 (.6407)	.9043 (.0999)

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 stationary distribution for y_{i1} . Results are reported for two-step GMM estimators.

Source: Blundell and Bond (1998), Table 2(a).

values for α of 0.5, 0.8 and 0.9. In all cases reported here, $\sigma_\eta^2 = \sigma_v^2 = 1$ and the initial conditions y_{i1} satisfy covariance stationarity. These results illustrate the poor performance of the first-differenced GMM estimator (DIF) at high values of α . Table I column 'DIF' presents the mean and standard deviation for this estimator in the Monte Carlo simulations. Consider the experiments where α is

0.8 or 0.9. For the first-differenced GMM estimator we find both a huge downward bias and very imprecise estimates. This is consistent with our analysis of weak instruments. For this reason, we consider further restrictions on the model which may yield more informative moment conditions.

3.2. Levels

If we are willing to assume that $E[\Delta n_{it}\eta_i^*] = E[\Delta k_{it}\eta_i^*] = 0$ and that the initial conditions satisfy $E[\Delta y_{i2}\eta_i^*] = 0$, then we obtain the additional moment conditions

$$E[\Delta x_{i,t-s}(\eta_i^* + w_{it})] = 0 \quad (3.5)$$

for $s = 1$ when $w_{it} \sim MA(0)$, and for $s = 2$ when $w_{it} \sim MA(1)$.¹ This allows the use of suitably lagged first differences of the variables as instruments for the equations in levels (cf. Arellano and Bover, 1995). Both sets of moment conditions can be exploited as a linear GMM estimator in a system containing both first-differenced and levels equations. Combining both sets of moment conditions provides what we label the system (SYS) GMM estimator.

For the AR(1) model, Table I shows that there can be dramatic reductions in finite sample bias from exploiting additional moment conditions of this type, in cases where the autoregressive parameter is only weakly identified from the first-differenced equations. This can also result in substantial improvements in precision. In contrast to the DIF estimator, there is virtually no bias and much better precision, even in the smaller sample size and for α of order 0.8.

3.3. Validity of the levels restrictions

To consider when these additional assumptions are likely to be valid in a multivariate context, we briefly consider the model

$$y_{it} = \alpha y_{i,t-1} + \beta x_{it} + (\eta_i + e_{it}) \quad |\alpha| < 1 \quad (3.6)$$

and close this by considering the following AR(1) process for the regressor

$$x_{it} = \gamma x_{i,t-1} + (\delta \eta_i + u_{it}) \quad |\gamma| < 1. \quad (3.7)$$

Thus $\delta \neq 0$ allows the level of x_{it} to be correlated with η_i , and we also allow

¹Further lagged differences can be shown to be redundant here if all available moment conditions for the equations in first differences are exploited. See Ahn and Schmidt (1995) and Arellano and Bover (1995).

$E[u_{it}e_{it}] \neq 0$. Both disturbances e_{it} and u_{it} are here assumed to be serially uncorrelated, and uncorrelated with the individual-specific effects η_i .

First notice that by repeated substitution after first-differencing (3.7), we can write

$$\Delta x_{it} = \gamma^{t-2} \Delta x_{i2} + \sum_{s=0}^{t-3} \gamma^s \Delta u_{i,t-s} \quad (3.8)$$

so that Δx_{it} will be correlated with η_i if and only if Δx_{i2} is correlated with η_i . To guarantee $E[\Delta x_{i2}\eta_i] = 0$ we require the initial conditions restriction

$$E\left[\left(x_{i1} - \frac{\delta\eta_i}{1-\gamma}\right)\delta\eta_i\right] = 0 \quad (3.9)$$

which would be satisfied under covariance stationarity of the x_{it} process.

Given this restriction, writing Δy_{it} similarly as

$$\Delta y_{it} = \alpha^{t-2} \Delta y_{i2} + \sum_{s=0}^{t-3} \alpha^s (\beta \Delta x_{i,t-s} + \Delta e_{i,t-s}) \quad (3.10)$$

shows that Δy_{it} will be correlated with η_i if and only if Δy_{i2} is correlated with η_i . To guarantee $E[\Delta y_{i2}\eta_i] = 0$ we then require the similar initial conditions restriction

$$E\left[\left(y_{i1} - \frac{\beta\left(\frac{\delta\eta_i}{1-\gamma}\right) + \eta_i}{1-\alpha}\right)\eta_i\right] = 0 \quad (3.11)$$

which would again be satisfied under covariance stationarity. Thus joint stationarity of the y_{it} and x_{it} processes is sufficient (but not necessary) for the validity of the additional moment restrictions for the equations in levels.

In our production function application, the moment conditions (3.5) would therefore be valid if the first moments of (n_{it}, k_{it}, y_{it}) were time-invariant (conditional on the common year dummies). However, this condition of jointly stationary means is stronger than required. For example, if the conditional model (3.6) has generated the y_{it} series for sufficiently long prior to our sample period for any influence of the true initial conditions to be negligible, an expression analogous to (3.10) shows that if Δx_{is} is uncorrelated with η_i in all periods, then Δy_{it} will be uncorrelated with η_i , even if the means of x_{it} and hence y_{it} are time-varying.

3.4. Testing the validity of the levels restrictions

As these additional moment conditions are overidentifying restrictions, their validity in a particular application can be tested using standard GMM tests of

overidentifying restrictions. In particular, a Sargan/Hansen test statistic can be based on the minimised value of the GMM criterion for the system GMM estimator. Moreover, since the moment conditions used by the first-differenced GMM estimator are a strict subset of those used by the system GMM estimator, a Difference Sargan test based on difference between the two standard Sargan statistics provides a more specific test of the additional moment conditions exploited by the system GMM estimator. Tests based on the minimised values of the two-step GMM criteria will have convenient chi-squared asymptotic distributions.²

One potential concern is that the small sample bias found for the first-differenced GMM estimator in the presence of weak instruments may give rise to serious size distortions in the behaviour of the associated Sargan test, and hence also in the Difference Sargan test. To investigate this, Table II reports the rejection frequencies obtained for these test statistics, using their asymptotic critical values at nominal 10%, 5% and 1% levels, in the Monte Carlo experiments of Table I. We do observe some tendency for these test statistics to reject a valid null hypothesis too often in these experiments, and this tendency is greater at higher values of the autoregressive parameter α . However these size distortions are not great, and they do not appear to be driven by particularly unreliable behaviour of the Sargan statistic for the first-differenced estimator. We also note that negative values were obtained for the Difference Sargan statistic in over 10% of the replications at $\alpha = 0.9$.³

A second potential concern is whether these tests have power to detect invalid moment conditions in this context. For example, Sowell (1996) shows that the standard Sargan test has no power against certain violations of moment conditions in a time-series setting. However simulations reported in Blundell and Bond (1998) suggest that the Sargan tests do have power to detect invalidity of the additional moment conditions used in the levels equations of the system GMM estimator, at least in the context of non-stationary initial conditions in an AR(1) model.

4. Data

The data we use is a balanced panel of 509 R&D-performing US manufacturing companies observed for 8 years, 1982-89. This data was kindly made available

²See Arellano and Bond (1991) for a discussion of these test procedures in the context of dynamic panel data models.

³Negative values were treated as not rejecting the null hypothesis in the reported rejection frequencies.

TABLE II
Rejection Frequencies

a) Sargan test: first-differenced GMM

N	α	10%	5%	1%
100	0.5	11.0	6.2	1.6
	0.8	13.1	6.8	1.7
	0.9	10.1	5.2	1.2
500	0.5	10.5	5.7	1.0
	0.8	11.1	7.0	1.2
	0.9	11.9	7.0	1.4

b) Sargan test: system GMM

N	α	10%	5%	1%
100	0.5	10.6	5.0	0.6
	0.8	12.7	6.3	1.2
	0.9	12.8	5.8	1.3
500	0.5	10.9	4.6	0.8
	0.8	10.5	6.0	1.0
	0.9	13.3	8.1	2.2

c) Difference Sargan test

N	α	10%	5%	1%	% negative
100	0.5	9.2	4.9	0.9	5.8
	0.8	14.4	6.7	1.5	11.5
	0.9	15.3	8.5	2.0	12.0
500	0.5	9.2	5.8	1.1	1.1
	0.8	11.8	6.4	0.7	6.2
	0.9	16.1	8.8	2.5	10.7

to us by Bronwyn Hall, and is similar to that used in Mairesse and Hall (1996), although the sample of 509 firms used here is larger than the final sample of 442 firms 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).

5. Results

5.1. Basic production function estimates

Table III reports results for the basic production function, not imposing constant returns to scale, for a range of estimators. We report results for both the unrestricted model (2.3) and the restricted model (2.1), where the common factor restrictions are tested and imposed using minimum distance.⁴ We report results for a one-step GMM estimator,⁵ for which inference based on the asymptotic variance matrix has been found to be more reliable than for the (asymptotically) more efficient two-step estimator. Simulations suggest that the loss in precision that results from not using the optimal weight matrix is unlikely to be large (cf. Blundell and Bond, 1998).

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, we reject the hypothesis that $\rho = 1$, and even using within groups we reject the hypothesis that $\rho = 0$. 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.⁶

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.⁷

⁴The unrestricted results are computed using DPD98 for GAUSS (see Arellano and Bond, 1998). The common factor restrictions are imposed using GAUSS, using a linearised minimum distance procedure described in Blundell, Bond and Meghir (1996).

⁵The one-step GMM estimator uses a weight matrix that does not depend on unknown parameters or residuals. For the first-differenced GMM estimator, we use the weight matrix discussed in Arellano and Bond (1991). For the system GMM estimator, we use the weight matrix discussed in Blundell and Bond (1998).

⁶The table reports p-values from minimum distance tests of the common factor restrictions and Wald tests of the constant returns to scale restrictions.

⁷p-values reported.

TABLE III
Production Function Estimates

	OLS Levels	Within groups	DIF t-2	DIF t-3	SYS t-2	SYS t-3
n_t	0.479 (.029)	0.488 (.030)	0.513 (.089)	0.499 (.101)	0.629 (.106)	0.472 (.112)
n_{t-1}	-0.423 (.031)	-0.023 (.034)	0.073 (.093)	-0.147 (.113)	-0.092 (.108)	-0.278 (.120)
k_t	0.235 (.035)	0.177 (.034)	0.132 (.118)	0.194 (.154)	0.361 (.129)	0.398 (.152)
k_{t-1}	-0.212 (.035)	-0.131 (.025)	-0.207 (.095)	-0.105 (.110)	-0.326 (.104)	-0.209 (.119)
y_{t-1}	0.922 (.011)	0.404 (.029)	0.326 (.052)	0.426 (.079)	0.462 (.051)	0.602 (.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 Sargan	-	-	-	-	.001	.102
β_n	0.538 (.025)	0.488 (.030)	0.583 (.085)	0.515 (.099)	0.773 (.093)	0.479 (.098)
β_k	0.266 (.032)	0.199 (.033)	0.062 (.079)	0.225 (.126)	0.231 (.075)	0.492 (.074)
ρ	0.964 (.006)	0.512 (.022)	0.377 (.049)	0.448 (.073)	0.509 (.048)	0.565 (.078)
Comfac	.000	.000	.014	.711	.012	.772
CRS	.000	.000	.000	.006	.922	.641

Asymptotic standard errors in parentheses.

Year dummies included in all models.

This is consistent with the presence of measurement errors. Instruments dated $t-3$ (and earlier) are accepted, 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. We expect this coefficient to be biased downwards if the instruments available are weak (cf. Blundell and Bond (1998) and Table I). Indeed the differenced GMM parameter estimates are all very close to the within groups results. 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 this is partly reflecting the increased power of the Sargan test to reject the instruments used in the first-differenced equations. A 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

Notes to TABLES III, IV and VI

$m1$ and $m2$ are tests for first-order and second-order serial correlation, asymptotically $N(0,1)$. We test the levels residuals for OLS levels, and the first-differenced residuals in all other columns.

Sargan is a test of the over-identifying restrictions for the GMM estimators. P-values are reported.

Dif Sargan is a test of the additional moment conditions used in the system GMM estimators relative to the corresponding first-differenced GMM estimators. P-values are reported.

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

GMM estimate. The constant returns to scale restriction is easily accepted in the system GMM results.⁸

5.2. Diagnosis

If the system GMM results are to be our preferred parameter estimates, we have to explain why the differenced GMM results should be biased. If the instruments used in the first-differenced estimator are weak, then the differenced GMM results are expected to be biased *in the direction of within groups*. Note that the first-differenced (one-step) GMM estimator coincides with a 2SLS estimator, exploiting the same moment conditions, when the firm-specific effects are eliminated using the orthogonal deviations transformation, rather than taking first-differences (Arellano and Bover, 1995). Note also that OLS in the model transformed to orthogonal deviations coincides with within groups (Arellano and Bover, 1995), and that weak instruments will bias 2SLS in the direction of OLS (Nelson and Startz, 1990a,b). Hence weak instruments will bias this particular 2SLS estimator (which coincides with first-differenced GMM) in the direction of within groups. Thus the similarity between our differenced GMM and within groups results suggests that weak-instruments biases may be important here.

To investigate this further, Table IV 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 is found to have an exact unit root. 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 found in Blundell and Bond (1998) and illustrated in Table 1. Indeed the surprise is that differenced GMM gives reasonable results for the employment series. One difference is that the variance of the firm-specific

⁸One puzzle is that we find little evidence of second-order serial correlation in the first-differenced residuals (i.e. an $MA(1)$ component in the error term in levels), although the use of instruments dated $t-2$ is strongly rejected. It may be that the e_{it} productivity shocks are also $MA(1)$, in a way that happens to offset the appearance of serial correlation that would otherwise result from measurement errors.

TABLE IV
AR(1) Model Estimates

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

Asymptotic standard errors in parentheses.

Year dummies included in all models.

TABLE V
Reduced Form Equations for 1989

Labour	First Differences	Levels
Wald	.021	.082
R ²	.032	.019
Capital		
Wald	.088	.001
R ²	.014	.034
Sales		
Wald	.631	.536
R ²	.011	.009

First Differences: Reduced Form regression of Δx_{t-1} on $x_{t-3}, x_{t-4}, \dots, x_{t-7}$.

Levels: Reduced Form regression of x_{t-1} on $\Delta x_{t-2}, \Delta x_{t-3}, \dots, \Delta x_{t-6}$.

Wald: p-value testing H_0 : slope coefficients jointly zero.

R²: coefficient of determination.

effects, relative to the variance of transitory shocks, is found to be lower 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 V reports some properties of the reduced form regressions from the AR(1) models. We focus on the 1989 cross section, where the largest set of lagged instruments is available. The reduced form regression for the first-differenced estimator relates $\Delta x_{i,88}$ to $x_{i,86}$ and further lags. These instruments are jointly significant in the employment reduced form, but not for capital or sales. This helps to explain why the differenced GMM estimator performs poorly in the models for

TABLE VI
Production Function Estimates imposing CRS

	OLS Levels	Within Groups	DIF t-3	SYS t-3
n_t	0.615 (.034)	0.632 (.034)	0.527 (.102)	0.516 (.101)
n_{t-1}	-0.559 (.035)	-0.135 (.033)	-0.204 (.109)	-0.288 (.117)
y_{t-1}	0.906 (.011)	0.361 (.029)	0.460 (.074)	0.562 (.079)
m1	-5.01	-9.78	-6.14	-6.80
m2	-2.36	-0.81	-0.61	-0.36
Sargan	-	-	.060	.065
Dif Sargan	-	-	-	.329
β_n	0.605 (.025)	0.655 (.034)	0.534 (.102)	0.518 (.079)
β_k	0.395 -	0.345 -	0.466 -	0.482 -
ρ	0.907 (.011)	0.424 (.025)	0.478 (.065)	0.563 (.068)
Comfac	.651	.000	.600	.966

Asymptotic standard errors in parentheses.

Year dummies included in all models.

capital and sales. The reduced form regression for the levels equations relates $x_{i,88}$ to $\Delta x_{i,87}$ and further lags. These instruments are jointly significant in the capital reduced form, although not for sales. This helps to explain why the system GMM estimator, which exploits both sets of moment conditions, works well for the capital series.

These results suggest that weak instruments biases are a potential problem when relying on first-differenced GMM estimators using these persistent series. This does not necessarily imply that weak instruments will be a problem when estimating the production function, since it may be that lagged combinations of the three series will be more informative than the lagged levels of any one series alone. However our results in Table III suggest that there may be important finite sample biases affecting the differenced GMM estimates of the production function. Moreover it is no surprise that the largest biases appear to be found on the coefficients for capital and lagged sales.

5.3. Constant returns to scale

Our preferred system GMM results in Table III accept the validity of the constant returns to scale restriction. Table VI considers imposing this restriction using each of the estimators. Two points are noteworthy. First, the validity of the moment conditions used to obtain the system GMM estimates becomes less marginal after imposing constant returns to scale. However the parameter estimates are very close to those found in Table III, and the common factor restriction continues to hold.

Second, the first-differenced GMM estimates are now much closer to the system GMM results, and not so close to the within groups estimates. Imposing constant returns to scale here seems to reduce the weak instruments biases in the differenced GMM estimates, possibly because the *capital-labour ratio* is less persistent than the levels of either series. This may provide some justification for the practice of imposing constant returns to scale in order to obtain reasonable estimates of the coefficient on capital, even though the restriction tends to be rejected with first-differenced estimators.

Both these points increase our confidence that the system GMM estimator works well in this application.

6. Conclusions

In this paper we have considered the estimation of a simple Cobb-Douglas production function using an 8 year panel for 509 R&D-performing US manufacturing companies. Our findings suggest the importance of finite-sample biases due to weak instruments when the first-differenced GMM estimator is used, although

these biases appear less important when constant returns to scale is imposed. We obtain much more reasonable results using the system GMM estimator: specifically we find a higher and strongly significant capital coefficient, and we do not reject constant returns to scale. We find that the additional instruments used in the system GMM estimator are both valid and informative in this context.

Whilst it would be dangerous to generalise from this one application, we can also report encouraging results from other applications of the system GMM approach. This has been applied to production function data for Britain and Germany (Bond, Harhoff and Van Reenen, 1998), to labour demand equations (Blundell and Bond, 1998), to investment equations (Bond, Harhoff and Van Reenen, 1999) and to cross-country growth regressions (Bond, Hoeffler and Temple, 1998). In each context the additional moment restrictions exploited by the system GMM estimator appear to be valid, and they appear to be useful in reducing finite-sample biases associated with first-differenced GMM.

ACKNOWLEDGMENTS

We thank Bronwyn Hall for kindly making available the data used in this paper. We acknowledge helpful comments from Manuel Arellano, Zvi Griliches, Bronwyn Hall, Dietmar Harhoff, Tor Jakob Klette, Jacques Mairesse, Laszlo Matyas, Frank Windmeijer, two anonymous referees and participants at the Göteborg meeting and in seminars at NIESR, Oslo and Bergen. This work forms part of the research programme of the ESRC Centre for the Microeconomic Analysis of Fiscal Policy at the Institute for Fiscal Studies.

REFERENCES

- Ahn, S.C. and P. Schmidt (1995), Efficient Estimation of Models for Dynamic Panel Data, *Journal of Econometrics*, 68, 5-28.
- Arellano, M. and S.R. Bond (1991), Some Tests of Specification for Panel Data: Monte Carlo Evidence and an Application to Employment Equations, *Review of Economic Studies*, 58, 277-297.
- Arellano, M. and S.R. Bond (1998), Dynamic Panel Data Estimation using DPD98 for GAUSS, mimeo, Institute for Fiscal Studies, London.
- Arellano, M. and O. Bover (1995), Another Look at the Instrumental-Variable Estimation of Error-Components Models, *Journal of Econometrics*, 68, 29-52.
- Blundell, R.W. and S.R. Bond (1998), Initial Conditions and Moment Restrictions in Dynamic Panel Data Models, *Journal of Econometrics*, 87, 115-143.

- Blundell, R.W., S.R. Bond and C. Meghir (1996), *Econometric Models of Company Investment*, in L. Matyas and P. Sevestre (eds.), *The Econometrics of Panel Data* (2nd edition), Kluwer Academic Publishers.
- Bond, S.R., D. Harhoff and J. Van Reenen (1998), *R&D and Productivity in Germany and the United Kingdom*, mimeo, Institute for Fiscal Studies, London.
- Bond, S.R., D. Harhoff and J. Van Reenen (1999), *Investment, R&D and Financial Constraints in Britain and Germany*, Working Paper no. W99/5, Institute for Fiscal Studies, London.
- Bond, S.R., A. Hoeffler and J. Temple (1998), *GMM Estimation of Empirical Growth Models*, mimeo, Nuffield College, Oxford.
- Griliches, Z. and J. Mairesse (1998), *Production Functions: the Search for Identification*, in S. Strom (ed.), *Essays in Honour of Ragnar Frisch*, Econometric Society Monograph Series, Cambridge University Press, Cambridge.
- Mairesse, J. and B.H. Hall (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. and B. Van Ark (eds.), *International Productivity Differences and Their Explanations*, Elsevier Science, 285-315.
- Nelson, C.R. and R. Startz (1990a), *Some Further Results on the Exact Small Sample Properties of the Instrumental Variable Estimator*, *Econometrica*, 58, 967-976.
- Nelson, C.R. and R. Startz (1990b), *The Distribution of the Instrumental Variable Estimator and its t-ratio When the Instrument is a Poor One*, *Journal of Business Economics and Statistics*, 63, 5125-5140.
- Sowell, F. (1996), *Test for Violations of Moment Conditions*, Discussion Paper, Graduate School of Administration, Carnegie Mellon University, Pittsburgh.
- Staiger, D. and J.H. Stock (1997), *Instrumental Variables Regression with Weak Instruments*, *Econometrica*, 65, 557-586.