NBER WORKING PAPER SERIES

ESTIMATING THE PRODUCTIVITY OF RESEARCH AND DEVELOPMENT: AN EXPLORATION OF GMM METHODS USING DATA ON FRENCH AND UNITED STATES MANUFACTURING FIRMS

Jacques Mairesse
Bronwyn H. Hall

Working Paper 5501

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
March 1996

This paper is part of NBER's research program in Productivity. Any opinions expressed are those of the authors and not those of the National Bureau of Economic Research.

NBER Working Paper 5501
March 1996

ESTIMATING THE PRODUCTIVITY OF RESEARCH AND DEVELOPMENT: AN EXPLORATION OF GMM METHODS USING DATA ON FRENCH AND UNITED STATES MANUFACTURING FIRMS

ABSTRACT

A comparative study of the contribution of R&D to firm-level productivity in French and United States manufacturing firms in the 1980s is presented. The study uses two large panels of approximately 1000 manufacturing firms covering over half of all R&D spending in each country and focuses on the estimation and interpretation of the relationship between output growth and the growth of R&D investment in the presence of simultaneity and firm heterogeneity. We use GMM methods to control for both sources of estimation bias, and we find 1) overall, the contribution of R&D to sales productivity growth appears to have declined during the 1980s, and 2) the role of simultaneity bias is higher in the U.S. than in France, possibly reflecting the greater importance of liquidity constraints for R&D investment in that country.

Jacques Mairesse
Institut National de la Statistique
et Etudes Economiques (INSEE)
CREST
15, Boulevard Gabriel Peri
92245 Malakoff Cedex
FRANCE
and NBER

Bronwyn H. Hall
Department of Economics
549 Evans Hall #3880
University of California, Berkeley
Berkeley, CA 94720-3880
and NBER

Estimating the Productivity of Research and Development in French and United States Manufacturing Firms: An Exploration of Simultaneity Issues with GMM Methods

J. Mairesse[*] *and B.H. Hall*[**]

[*] *Centre de Recherche en Economie et Statistique and National Bureau of Economic Research*

[**] *University of California, Berkeley and National Bureau of Economic Research*

1. Introduction

Industrial firms in developed economies engage in increasing amounts of organized research and development activity aimed at producing new and improved products and reducing costs. Economists would like to know the answers to two questions concerning the success (or failure) of this R&D activity: First, what is the magnitude of the returns earned by the firms that undertake it? Do these returns justify the investment being undertaken? Second, to what extent do the benefits of the R&D spill over to other firms, thus lowering their innovation costs, or to the firm's customers, through lower prices and improved products? In brief, what are the private and social returns to the R&D being performed. The standard approach to answering these questions is grounded in the productivity residual methodology: we measure the contribution of R&D to a firm's revenue or quantity of output, controlling for the other inputs into production. The former measure (the marginal revenue elasticity) is the relevant concept for the computation of private returns: it is the sum of the contribution of R&D to the equilibrium price of quality-adjusted goods sold by the firm and the contribution of R&D to the equilibrium quantity of quality-adjusted goods sold. The latter measure (the output elasticity) is what matters more for society as a whole: how the gains in productivity get allocated between the firm and its customers is of secondary importance for growth, although the allocation does affect the firm's incentive to undertake R&D.

One way to enrich our understanding of the sources and causes of productivity levels and growth is by using cross-country, cross-industry, and cross-time comparisons in order to isolate those features that are robust to changes in time, place, and institutions. The differences or variations that emerge from studies of this kind are also informative, particularly when they are linked to other known differences in the economic environment. A large number of studies of the relationship between R&D investment and productivity growth at the firm level have been conducted in the past using data through 1980 (see Mairesse and Sassenou, 1991, for a survey), but little has been done with the increasing amount of data which became available during the 1980s in most of the "big seven" OECD countries. This paper uses new datasets available in both the U.S. (a dataset based on Compustat files that has been updated to 1990 and then merged with deflators at the two-digit industry level)[1] and France (an R&D survey at the firm level, also updated to 1990, merged with conventional enterprise data on production)[2] to investigate whether the small but well-documented relationship between R&D and productivity growth at the firm level persisted during the 1980s and whether it is similar in the USA and France. While so doing, we distinguish somewhat more carefully than past studies between revenue growth and productivity growth itself.

The style of analysis is based on the traditional growth framework and draws on our experience with analyzing the French data for 1980-1987, which is described in Hall and Mairesse (1992). Since the work described in that paper, we have obtained a larger French sample and somewhat improved our deflators (we are now using sales and value added deflators at the two-digit level).

The US data is of somewhat lower quality than the French data, particularly since we do not have a measure of value added, nor do we have a reliable measure of labour costs. In addition, we do not have the information necessary to correct the capital and labour measures for R&D double counting as we do in France. Here, comparison of results using the French data can help. From these data we are able to gauge the changes which result when better measures of both right-hand-side and left-hand-side variables are used. The evidence suggests that estimates using sales instead of value added are not too badly biased, but that attempting to correct for materials or including materials directly in the regression can give misleading results. Correcting for double counting of R&D employees tends to rais the R&D significantly, which is consistent with an earlier work.

The plan of the paper is as follows: we first describe the data samples which we are using, and characterise the overall similarities and differences of the manufacturing sectors in the two countries. This is followed by a detailed examination of the form of the productivity-R&D relationship using the French data, where we have better variables. Once we have chosen a specification which is feasible for the US data as well, we present comparative results for the two countries. Since we find that the dating of our capital stocks (physical and R&D)

---

[1] See Hall (1990) for a description of a slightly earlier version of the US data used here.

[2] The two sources are the *Enquête annuelle sur les moyens consacrés à la recherche et au développement dans les entreprises*, conducted by the French Ministry of Research and Technology, and the *Enquête annuelle des entreprises*, conducted by INSEE.

affects the within-firm estimates greatly (with end of period capitals having higher coefficients than beginning period), the final section of the paper explores the role of simultaneity in the relationship, and presents in details GMM estimates of the relationship that are more efficient than conventional first differenced estimators with lagged right-hand-side variables as instruments.

2. Samples, Framework, and Variables

Table 1 shows some of the characteristics of the samples with which we will be working. In each case, we began with an unbalanced panel from 1981 to 1989, with up to 3 years of lagged values for each variable (that is, the actual data set goes from 1978 to 1989, and no firm has less than 3 years of data). Later in the paper we use two fully balanced subsets of data for each country. These subsets contain data for shorter periods, 1978 to 1985, and 1982 to 1989.[3] The data have been cleaned so that there are no jumps in the stock variables of absolute value greater than 200 percent, or in the flow variables of absolute value greater than 300 percent. Both samples cover a large fraction of the relevant population: the US sample has 50 percent of manufacturing employment and about 67 percent of industrial R&D in 1985,[4] whereas the French sample has about 22 percent of manufacturing employment and 56 percent of industrial R&D.[5] In both cases, firms had to perform and report R&D during the period to be in the sample, so there is some selectivity at work.

The samples for the two countries are fairly similar in industrial distribution. The most striking differences are the large number of computing, electronics and instrument firms in the United States, and pharmaceutical, chemical, food and machinery firms in France. The balanced samples used in the latter half of the paper omit a large number of the Computing and Electronics firms in the USA, many of whom are small recent entrants, and are somewhat more heavily weighted toward the food and pharmaceutical industries in France.[6]

---

[3] For the USA, the unbalanced panel has 1073 firms and the balanced subpanels 535 and 442 firms respectively, whereas for France, the numbers are 1232, 447, and 381. There is a substantial overlap between the two different subperiods in each country; this overlap is greater for France than for the USA.

[4] According to the National Science Foundation, domestic R&D expenditure in 1985 was 58 billion dollars, while our sample had total R&D spending of 39 billion dollars.

[5] According to the OECD, total R&D performed by business enterprises in France in 1985 was $6.04 billion dollars (using a purchasing power parity rate of 7.27 francs per dollar to convert from francs to dollars), while our sample has $3.37 billion dollars of R&D.

[6] In both countries the aircraft and other transportation sector has an extremely high R&D-to-sales ratio, and it has about 50 percent of the private enterprise-performed, government-funded R&D. Because we believe that estimating the productivity of R&D in this sector may be problematic due to fact that a primary customer is the go-

TABLE 1
Unbalanced Sample Characteristics: French and US Manufacturing, 1981-1989

Industry	Number of Firms	Number of Observations	Employment (000s)	R&D-Sales Ratio
		United States		
Electronics, Computers, & Inst.	382	2254	2209	7.06
Pharmaceuticals	100	623	761	5.56
Chemicals	34	263	730	3.73
Autos	43	266	1668	3.23
Electrical Machinery	66	377	565	3.32
Machinery	135	899	832	2.59
Rubber & Plastics	36	204	279	2.16
Paper & Printing	41	264	395	2.02
Fabricated Metals	62	380	260	1.54
Wood, SCG, & Misc.	71	426	314	1.17
Primary Metals	27	162	195	1.12
Textiles & Leather	39	190	167	0.90
Food	37	213	879	0.95
Total	1073	6521	9254	2.93
Aircraft & other trans.	26	165	1074	3.60
		France		
Electronics, Computers, & Inst.	186	910	188.6	6.04
Pharmaceuticals	191	1081	83.4	2.92
Chemicals	119	540	65.0	1.54
Autos	62	312	247.8	1.48
Electrical Machinery	109	538	94.6	2.24
Machinery	192	933	73.6	1.17
Rubber & Plastics	62	322	63.9	2.69
Paper & Printing	32	123	10.2	0.49
Fabricated Metals	78	378	29.3	0.78
Wood, SCG, & Misc.	38	240	44.5	0.82
Primary Metals	39	201	41.9	0.50
Textiles & Leather	38	202	13.6	0.82
Food	86	502	65.1	0.29
Total	1232	6282	1021.7	2.26
Aircraft & other trans.	32	195	93.7	9.41

Employment and R&D to sales are for the middle year of the sample, 1985. The R&D to sales ratio shown is a *sales-weighted* average, which is the *industry* R&D to sales ratio.

The framework in which we measure the contribution of R&D to productivity growth is a standard growth accounting one, based on the Cobb-Douglas production function.[7] The basic equation is the following:

$$y_{it} = a_i + \lambda_t + \alpha c_{it} + \beta l_{it} + \gamma k_{it} + \varepsilon_{it} \tag{1}$$

where i and t index firms and years respectively, y is output, c is capital, l is labour, k is knowledge or R&D capital, and the lower case letters denote logarithms. The equation allows for both additive firm and year effects. In this formulation, y denotes a value added output concept, since materials are not included in the model. Although we have a measure of value added for the French data, we do not have one for the US data. Therefore we present estimates using both sales and value added, and also including materials for the French data, so that we can calibrate the results using sales for comparison to the US results.

Our measure of capital stock is a constructed estimate of plant and equipment adjusted for inflation in both countries. Our measure of R&D capital is that described as K71 in Hall and Mairesse (1995) for France and in Hall (1990) for the USA. In both cases it is constructed from the past history of R&D investment, with a depreciation rate of 15 percent per year and a pre-sample growth rate of 5 percent per year. Our measure of labour is the number of workers in the firm. This is usually reported by the firms as the average number of workers during the year. In the United States, it may occasionally be the number of workers at the close of the fiscal year.

Conceptually, the value added, labour, and capital measures used to estimate equation (1) should be purged of the contribution of R&D materials, physical capital used in R&D laboratories, and R&D personnel, since these inputs do not produce current output, but are used to increase the stock of R&D capital. If this is not done, the cross section estimates (or estimates from firms in long run equilibrium where R&D spending does not change much from year to year) will not necessarily be incorrect, but the measured R&D coefficient will be some kind of "excess" elasticity of output to R&D rather than a total elasticity, i.e. the incremental productivity of R&D above and beyond the normal productivity of the capital and labour involved. In Hall and Mairesse (1992) we confirmed this interpretation, finding that estimates corrected for R&D inputs tended to give R&D elasticities which were 0.06-0.08 higher than uncorrected estimates (and that most of the effect could be achieved by correcting the labour variables).

vernment, we have omitted it in the regressions that follow. In fact, removing it changed the results very little.

[7] For more details, see our earlier paper (Hall and Mairesse, 1995).

Here we correct only value added and labour for the French data, but we are unable to do so for the US data, since we do not have the appropriate measures.

3. Comparing the Conventional Estimates for France and the United States

Table 2 presents a series of estimates of equation (1) for the French data. The four horizontal panels have increasingly less restrictive assumptions on the error term. The first is a regression pooled across firms and time, with individual year dummies, while the second also allows for industry effects (at the sectoral level shown in table 1). The third and fourth allow for additive firm effects, first the estimator with overall firm means removed, and then estimates in growth rates. Except for simultaneity and measurement error bias in the right-hand-side variables, these last two panels should have the same estimates. On the whole, the two capital coefficients appear to be similarly insignificant, while the labour coefficient is somewhat lower for the growth rate estimates, suggesting the presence of measurement error.

The first column of the table shows the basic specification which we will also use for the US data. The second column displays the same sales regression with materials included on the right-hand-side, while the third uses value added instead of sales. The average materials share in these data is sixty percent, so the estimated coefficient is somewhat high, especially when permanent differences across firms are controlled for. This is typical of these kinds of data and can happen for two reasons: the measurement error bias can be less for materials than for other inputs[8], or there are shortrun increasing returns to materials.[9]

How do the estimates using sales in column (1) compare with those which either include materials, or use value added as a dependent variable? If we simply compare column (1) with column (3), we can see that the labour coefficient is typically somewhat lower for sales, while the capital coefficient is somewhat higher in the cross section dimension, but about the same and insignificant in the within dimension. The estimates using sales and excluding materials seem to give results for R&D capital that are quite similar to those using value added. Unfortunately, they are also insignificant in the within-firm dimension. Adding materials to the equation merely reduces the coefficients by the estimated materials share, but their magnitudes are more or less what one would predict from the value added equation. The conclusion is that the regression of sales on labour, capital, and knowledge capital is likely to give results which are quite similar to those obtained using value added as a dependent va-

[8] Griliches and Hausman (1986).

[9] Shortrun increasing returns can occur for any of a number of reasons, most, but not all, of them involving a failure of perfectly competitive conditions. See R. E. Hall (1988) for further discussion of this point.

TABLE 2
Productivity Regressions 1981-1989, France (6282 observations)

Capital Dating	Beginning of Year				End of Year
Dep. Variable	Log Sales	Log Sales	Log VA	Log VAC	Log VAC
Total					
Log L	.591(.017)	.193(.005)	.699(.012)	.630(.012)	.597(.012)
Log C	.295(.012)	.043(.002)	.193(.008)	.183(.008)	.210(.009)
Log K	.090(.006)	.024(.001)	.092(.004)	.165(.004)	.172(.004)
Log M		.735(.004)			
R^2(s.e.)	.868(.489)	.993(.115)	.926(.349)	.923(.357)	.927(.347)
Within Ind.					
Log L	.681(.011)	.201(.003)	.749(.008)	.679(.008)	.645(.008)
Log C	.204(.009)	.038(.002)	.153(.007)	.141(.007)	.168(.007)
Log K	.109(.006)	.023(.002)	.093(.004)	.167(.004)	.176(.004)
Log M		.734(.003)			
R^2(s.e.)	.899(.429)	.993(.112)	.933(.333)	.930(.340)	.933(.332)
Within Firm					
Log L	.819(.013)	.199(.006)	.900(.017)	.841(.016)	.790(.017)
Log C	-.045(.013)	.001(.005)	-.036(.016)	-.007(.015)	.050(.016)
Log K	.008(.011)	-.010(.004)	-.016(.013)	.013(.013)	.069(.014)
Log M		.791(.005)			
R^2(s.e.)	.713(.143)	.956(.056)	.597(.178)	.602(.176)	.606(.175)
First Diff.					
Log L	.645(.032)	.154(.012)	.715(.035)	.666(.032)	.606(.032)
Log C	-.001(.007)	-.002(.002)	-.006(.008)	-.003(.008)	.130(.025)
Log K	-.003(.003)	.000(.001)	-.005(.003)	-.004(.003)	.080(.021)
Log M		.793(.011)			
R^2(s.e.)	.256(.146)	.878(.059)	.192(.185)	.182(.187)	.190(.186)

All equations contain year dummies. The industry dummies used in the second panel are at the level given in table 1. Variables:

Log L	Logarithm of average employment during the year.
Log C	Logarithm of gross plant and equipment at the beginning or end of the year, adjusted for inflation.
Log K	Logarithm of the R&D capital stock at the beginning or end of the year, as computed in Hall and Mairesse (1995).
Log M	Logarithm of materials expenditures during the year.
Log Sales	Logarithm of sales during the year deflated by an overall manufacturing deflator.
Log VA	Logarithm of value added during the year.
Log VAC	Logarithm of value added during the year, corrected for R&D material cost. In this column, Log L has also been corrected for the number of R&D employees.

riable (possibly with a slightly lower labour coefficient in all dimensions and a slightly higher physical capital coefficient in the totals).

The final two columns of table 2 investigate two questions: first, what is the effect of correcting labour and value added for R&D inputs, and second, what are the differences in the estimates when we use end of year capitals rather than beginning of year? The answer to the first question is that the double counting corrections raise the R&D capital elasticity by about .07 in the cross section dimension, .03 in the within dimension, and not at all in first differences. This is entirely consistent with our earlier results (which use a smaller sample for 1980 to 1987) and those of Cuneo and Mairesse (1984) (which use data from 1972 to 1977), as well as those of Schankerman (1981) (which use US data in the cross section dimension only).

Using end of year capital stocks rather than beginning of year raises the capital coefficients slightly in the cross section, but it changes the results dramatically in the time series dimension. The within-firm physical and R&D capital coefficients both triple, and the first difference estimates go from essentially zero to quite plausible numbers which are closer to the shares of both capitals in value added. Why does this happen? Because the capital used in production during the year is likely to be some weighted average of beginning and end of year capital, it is difficult to know precisely which dating to use, but either one ought to work about as well (or as badly) unless something other than a simple production relationship is driving the regression. Unfortunately, the most likely explanation is simultaneity between changes in value added (or sales) and investment of both types, driven either by demand shocks or liquidity shocks. This is why we explore the use of instrumental variables to correct for simultaneity bias and reexamine this question later in the paper.[10]

Our explorations with the French data give us confidence that there is information in the simple sales regression without materials inputs, which is all we can estimate using US data. We therefore present estimates of the sales productivity regression for the United States in table 3, together with estimates of the identical model for France (the first column repeated from table 2). We use two different measures of sales as the dependent variable: sales deflated by a single manufacturing sector deflator, and sales deflated by a two-digit level deflator. The French cross section estimates are more or less comparable with those of earlier periods, but the estimates within firm are lower: Cuneo and Mairesse (1984) obtain .11 for these estimates on a sample of 182 firms from 1972 to 1977 (where the data are corrected for double counting, but they also show that these corrections do not make much difference in within-firm estimates). When we use a single common sales deflator for all industries, the results for the

[10] For general discussion of simultaneity issues in estimating production functions, see Griliches and Mairesse (1995).

United States are slightly weaker: where Griliches and Mairesse (1984) obtained .05 in cross section and .09 within using 133 firms from 1966 to 1977, the results here are about .04 both in the cross section and within firms. They also agree with those of Hall (1993b), which were obtained using a superset of these firms.

In contrast to these fairly weak results, using sales deflators at the two-digit level (the columns labelled Log S (2-D)), raises the estimated R&D capital coefficient in the United States substantially, from about 0.04 to 0.25 in totals, from 0.04 to 0.17 in within firms, and from 0.01 to 0.09 in growth rates. For

TABLE 3
Productivity Regressions 1981-1989

Observations	United States (6521)		France (6282)	
Dependent Variable	Log Sales	Log S (2-D)	Log Sales	Log S (2-D)
Totals				
Log L	.666(.010)	.779(.021)	.591(.017)	.578(.017)
Log C	.289(.008)	-.021(.018)	.295(.013)	.303(.012)
Log K	.035(.005)	.246(.012)	.090(.006)	.093(.006)
R^2(s.e.)	.936(.360)	.816(.849)	.868(.489)	.862(.499)
Within Ind.				
Log L	.712(.009)	.707(.021)	.681(.011)	.681(.011)
Log C	.204(.008)	.128(.019)	.204(.009)	.204(.009)
Log K	.063(.005)	.173(.013)	.109(.006)	.110(.006)
R^2(s.e.)	.967(.339)	.837(.800)	.899(.429)	.898(.430)
Within Firm				
Log L	.742(.010)	.702(.013)	.819(.013)	.828(.014)
Log C	.126(.011)	.142(.014)	-.046(.013)	-.052(.013)
Log K	.041(.011)	.170(.014)	.008(.011)	.013(.011)
R^2(s.e.)	.720(.145)	.673(.193)	.713(.143)	.530(.145)
First Diff.				
Log L	.585(.019)	.592(.020)	.645(.032)	.648(.032)
Log C	.148(.020)	.171(.021)	-.001(.007)	.001(.007)
Log K	.010(.024)	.092(.026)	-.003(.003)	-.003(.003)
R^2(s.e.)	.416(.156)	.402(.167)	.256(.146)	.249(.146)

The variables are the same as those in table 2 with capital stocks at the beginning of the year. All equations contain year dummies. The industry dummies are at the level given in table 1. The standard error estimates for the total and first differenced estimates are heteroscedastic-consistent, those for the within industry and within firm estimates are conventional estimates.

the French data there is no such increase. The reason for this is quite simple: the United States deflator for the computing sector is based on a hedonic price index for computers, and as a consequence it declines by about 80 percent during the 1980s as computer hardware becomes much more powerful and much less expensive. Because this industry also has a growing R&D budget, this price decline and the attendant growth in real output means that the estimated contribution of R&D to the output productivity of this industry is very substantial.[11] However, in terms of sales undeflated or only deflated by our overall deflator, the firms in this industry have not seen the same kind of productivity gain, because consumers have captured most of the benefits in the form of cheaper computers. In France, on the other hand, the computing deflator is a more conventional measure that does not capture the tremendous decline in the price of raw computing power, so deflation at the industry level does not have the same effect on the estimated productivity of R&D.

A remaining mystery in our results is the within-firm physical capital coefficient for the French firms, which is negative, sometimes significantly so. It is unclear why this is so: it was true throughout table 2 also, except in the last column where we used end of year capital. This result contrasts with those of Cuneo and Mairesse (1984) for the seventies and our own earlier results for 1980-1987. Understanding this puzzle awaits future work with the French data.

4. Trying to Correct for Simultaneity Bias with GMM

The sensitivity of our estimates to the dating of capital stock suggests that the assumption of zero correlation between regressors and disturbances necessary for the consistency of estimates is unlikely to be justified. We have already indicated that this correlation can arise because of simultaneity between sales and both types of investment. The failure of the non-correlation assumption in panel data causes more problems than it does in conventional regression estimation, because it frequently invalidates estimates based on data where firm means have been removed, even when instrumental variable estimation is used to correct for simultaneity.[12] This is because the only instruments normally available are

[11] Some of this gain may be mismeasured or even overstated, because important inputs to the computer industry (semiconductors and computer components) have also suffered substantial price declines during the same period, but these have not been captured at all in our regression. These inputs are typically contained in the materials input to the firm's production at cost. When there are large price changes in these inputs, the assumption for the purpose of production function estimation that the cost of the omitted materials and intermediate goods is a constant fraction of sales within a firm, is more likely to fail.

[12] Firm means are removed by subtracting $\bar{y}_i = \beta\bar{x}_i + \alpha_i + \bar{\varepsilon}_i$ from the model in equation (2), leaving the equation to be estimated:

lagged values of the right hand side variables and in short panels the correlation of these variables with the disturbance remains after the firm-level means of the dependent variable have been removed. The usual solution to this problem is to use first differences (growth rates) for estimation rather than the within firm correction. In addition to simultaneity bias, another source of bias likely to be present is that arising from measurement error in all the variables; under a variety of assumptions, this error tends to bias the first differenced coefficients more towards zero than does the within-firm estimator (Griliches and Hausman, 1986). Thus it is desirable to develop within-firm estimators that remain consistent even when lagged instruments are used.

Since earlier work with these kinds of data (Griliches and Mairesse, 1984), a series of papers have been published which attempt to systematise the methods for estimating and testing the validity of instrumental variable estimates of panel data models where there may be simultaneity, measurement error, and effects correlated with the regressors (Griliches and Hausman, 1986; Arellano and Bond, 1991; Keene and Runkle, 1992; Schmidt, Ahn, and Wyhowski, 1992). With the exception of Griliches and Hausman, who take a some what different route, the approach followed in these papers is to set up a series of successively stronger (more numerous) orthogonality conditions which are valid under various versions of the panel data model and to use the results of Generalized Method of Moments (GMM) estimation on these conditions to choose among the specifications. The appeal of this method for panel data rests in the weakness of the distributional assumptions necessary to carry it out: it does not require the assumptions of zero covariance across years (except to the extent that this is necessary to validate the instruments), or homoscedasticity across firms for efficiency.[13] The standard error estimates which emerge from a GMM estimation are also robust to the presence of correlation across equations and heteroscedasticity. There is a cost to this flexibility of course, in the form of somewhat larger standard errors, but this may be of less concern when we are dealing with large panel datasets.

In this section of the paper, we use the GMM approach to panel data

---

$$y_{it} - \bar{y}_i = \beta\,(x_{it} - \bar{x}_i) + \varepsilon_{it} - \bar{\varepsilon}_i$$

This removes the firm effect α_i but contaminates the disturbance ε_{it} with the disturbances ε_{i1} , ..., ε_{iT} from the other years of data.

13 Allowing for heteroskedasticity when estimating with heterogeneous groups of firms is essential, if only because we cannot sustain the assumption that the production function we are estimating is identical across firms. If instead of $y_{it} = \beta x_{it}$, the model is $y_{it} = \beta_i x_{it}$, estimating an average β across firms will perforce introduce a size-related heteroskedasticity into the disturbance, of the form $(\bar{\beta} - \beta_i)x_{it}$. If the variation in β is random across firms, no bias will ensue, but the standard errors will be wrong unless we allow for this heteroskedasticity. If the variability is related to the average level of x_{it}, it will be absorbed in the fixed effect, and cause no bias.

estimation in order to investigate the importance of correlated effects, simultaneity, and measurement error for our estimates. As instruments for labour, capital, and R&D capital, we use these same variables lagged 3 years and we test for the admissibility of more recent lags as instruments. Although this choice of instruments will be justified if simultaneity is of concern, when the source of bias is measurement error in the capitals, this measurement error is likely to be correlated over time, and lagged capital not a very good instrument. We investigated this possibility using investment and R&D expenditures as instruments for the capital stocks but found that using these alternative instruments made little difference and that we were able to accept the validity of the capital stocks as instruments in the presence of investment and R&D expenditures. To save space, the results of this investigation are not reported here.

To make our approach clear, consider the simple regression model with a single regressor, but with panel data:[14]

$$y_{it} = \beta x_{it} + u_{it} = \beta x_{it} + \alpha_i + \varepsilon_{it} \qquad i = 1,\dots,N; \; t = \tau+1,\dots,T \tag{2}$$

where there are τ periods of data available as instruments for the first year of estimation. β is the parameter of interest, and α_i is the firm effect, potentially correlated with x_{it}, $t = 1,\dots,T$. Our maintained model is that the effects are correlated (so we have to difference the equation) and that only lag 3 and higher x's are available as instruments for x_{it}, because later values are correlated with ε_{it}.[15] These assumptions imply the following set of orthogonality conditions:

$$E\,[x_{is}\,\Delta u_{it}] = 0 \qquad t = \tau+2,\dots,T; \; s = 1,\dots,t-\tau-1 \tag{3}$$

where $\Delta u_{it} = \Delta y_{it} - \beta\Delta x_{it}$. There are $(T-\tau-1)(T-\tau)/2$ orthogonality conditions in (3).

Estimation of (3) is performed using the method described in the Appendix. Consistent estimates can be obtained simply by minimizing the sum of squares of the empirical moments $[f(\beta)]$ corresponding to (3) with respect to β, but efficiency requires that we also form an estimate of their covariance $\hat{\Omega}$ and

---

[14] Our presentation from now on will suppress the presence of λ_t, the time effect, in the model. In estimation, we remove the year means from the data at the very start to avoid complications. This procedure has no effect on either the consistency or efficiency of estimation in the case of a linear model with additive time dummies. In fact, GMM estimates with time dummies excluded and year means removed are numerically identical to those with time dummies included (and a vector of ones included in the instruments).

[15] This choice is based on prior experience with firm data (Hall, 1991; and Blundell, Bond, Devereux, and Schiantarelli, 1992). It errs on the side of caution and is, in fact, not strictly necessary in these data, as we shall see.

minimise in that metric. In large samples (asymptotically), the minimised statistic $f(\hat{\beta})\ \hat{\Omega}^{-1}\ f'\ (\hat{\beta})$ is distributed as a chi-squared random variable with degrees of freedom equal to the number of orthogonality conditions in (3) less the number of estimated parameters β. As an example, suppose T is 8 and τ is 3, as in our data; then the number of orthogonality conditions (3) is (8-3-1)(8-3)/2 = 10. In our equation (2) example there is one β, so that the degrees of freedom will be 9. It can be shown that the set of moment conditions in (3) is equivalent to that used in the instrumental variable estimation of the within estimator of (2) where the firm-level means have been computed only over observations which postdate any disturbance that might be correlated with the instruments. Both use all the available information in the data.[16]

We use (3) as our basic specification, with sales as the dependent variable and labour, capital and R&D capital as the independent variables, and then test for the additional moment restrictions implied by the validity of lag 2 instruments, lag 1 instruments, lag 0 instruments (weak exogeneity), and then by using all years as instruments for all equations (strong exogeneity). We then compare each of these specifications individually with specifications of the following form:

$$E\ [x_{is}\ u_{it}] = 0 \qquad t = \tau+1,\ldots,T;\ \ s = 1,\ldots,t-\tau-1 \tag{4}$$

The moment conditions defined by equation (4) are appropriate if the firm-level effects are not correlated with the x's. In order for the procedure outlined above to be valid and in order to guarantee a non-negative chi-squared statistic for the tests, it is necessary to use an equivalent consistent estimate of $\hat{\Omega}$ for all the estimates; we ensure this by forming our estimate of the sample covariance of the $\{x_{is}\Delta u_{it}\}$ or $\{x_{is}u_{it}\}$ using estimates of u_{it} that are based on the βs estimated using the weakest specification (that of equation (3) with only lag 3 and higher instruments).

A drawback to GMM estimation for panel data under the current state of the art is that the results of asymptotic theory are somewhat incomplete when the datasets are unbalanced. Conceptually, if there are a small number of patterns of missing data, it would be possible to estimate the model separately for each subset of firms which had data in a particular set of years, and then combine these estimates optimally using their variance estimates as weights (see Bound, Griliches, and Hall, 1986, for a discussion and demonstration of a similar methodology applied to maximum likelihood estimation). Provided the number

---

[16] Schmidt, Ahn, and Wyhowski show that there are additional moment restrictions involving second moments of the disturbances available for estimation when their variances do not change over time; we have not exploited these here because the assumption of constant variances does not appear to hold in our data.

of subsamples is fixed and the number of firms within each is allowed to grow, it would be possible to obtain asymptotic results for such a method. However, implementing this approach to estimation is extremely cumbersome, and we have chosen a simpler solution here. We created two subsets of the data that are balanced for subperiods within our overall period: the first is 1981-1985 (with the years 1978, 1979, and 1980 available as instruments) and the second is 1985-1989 (with 1982-1984 available as instruments).

The results of our sequence of specification tests using these balanced panels are shown in graphs 1 and 2 for France and graphs 3 and 4 for the United States as nested χ^2 graphs (see pages 19 to 22). Choosing the appropriate size for such a sequence of tests is of course somewhat arbitrary; we have selected the one per cent level of significance because of the fairly large number of tests conducted. In both periods (1981-1985 and 1985-1989), the absence of correlation between firm effects and the instruments is clearly rejected in all cases for the French and the US data, somewhat weakly for the United States in the first period. The US data reject weak exogeneity in both periods in favour of lag 1+ instruments. This can be due either to simultaneity bias or to measurement error in the right hand side variables. The picture for France is slightly more confused: in the first period, the data accept the strong exogeneity of the right hand side variables, while in the second strong exogeneity is rejected in favour of lag 0+ instruments (weak exogeneity). We chose to use the weaker set of instruments for both periods so that the results would be comparable.

In tables 4 and 5 we display the results of our preferred estimations using GMM with lag 0+ instruments for France and lag 1+ instruments for the United States (fifth column). For comparison, we also show the GMM estimates only using the lag 3+ instruments (fourth column), and the conventional total and first differenced estimates of the production function. The latter is estimated with both beginning and end of year capital stocks, and the former with beginning of year stocks (three first columns). Note that since we are trying to correct for simultaneity bias, our GMM estimates are performed on the specification with the end of year capital stocks. The pattern of the estimates are roughly similar across countries. Both countries have fairly large capital coefficients in the totals which are very substantially reduced in the differenced estimates, especially when beginning of year capital is used or capital is instrumented. As in the results for the whole period, the coefficient of R&D capital is somewhat larger in the totals for France than for the United States, but disappears or turns negative in the first differences in the two countries. One interpretation could be that firms which invest heavily in R&D have higher productivity growth on average in the long run, but there is weaker evidence of a short run effect. Note also that the physical capital coefficient in France has not been helped by obtaining estimates which correct for simultaneity; in the within-firm dimension, it is still implausibly low, even negative, albeit somewhat poorly measured. This is less true in the United States; correcting for simultaneity bias lowers the physical capital coefficient only slightly, while wiping out the R&D capital coefficient.

Despite the roughly similar pattern of our estimates, it is interesting to indicate a possibly substantial difference between the two countries which may help to explain why we were able to accept the weak exogeneity of the capital stocks for France, but not for the United States. As already said, the influence of current sales on investment, either because sales signals future demand or because of liquidity considerations, is a very likely source of simultaneity between end of year capital stocks and current year sales. In France, the correlations of the growth rate of sales with that of investment and R&D are 0.21 and 0.12 respectively, while for the United States these numbers are 0.30 and 0.32. Thus this particular source of simultaneity bias is potentially higher in the United States than in France. It is tempting to speculate whether this difference is due to a greater sensitivity to financial constraints in the US firms, especially because the sample is drawn entirely from publicly traded enterprises that may find it difficult (or expensive) to finance any investment, and particularly investment in intangibles such as R&D, in the public equity or debt markets. The French firms, on the other hand, may face somewhat softer budget constraints, and the heavier involvement of the government in industrial R&D may mitigate the effects of sales or earnings fluctuations. At the moment, however, this is just speculation.

Finally, the substantive result emerging from tables 4 and 5 is that the "excess" productivity of R&D is essentially zero in both periods in the United States and France (and even possibly slightly negative in France). The former result agrees with the finding in Hall (1993b) that, during the 1980s, the market value of the R&D investment undertaken by this sample of firms indicated that no excess returns were expected from such investment.

Looking now at our GMM estimates from an efficiency technical point of view, the bad news is that adding more moment conditions by using all the available instruments leaves us with standard errors of our estimated coefficients which are much larger than the (heteroskedastic consistent) standard errors of the usual first differenced estimates (reported in column 3 or column 2). This is particularly striking for the labour and ordinary capital coefficients, but much less for the R-D capital coefficient. The first stage R-squares suggest why: the growth of R&D capital is much more predictable from the past levels of capital and labour than is the growth of ordinary capital and the coefficient of this variable is therefore somewhat better measured than the others when we instrument it.[17]

[17] These R-squares are the following: for France, .073, .295, and .531 for the growth

TABLE 4

Estimates for the Balanced French Panels, Dependent Variable: Log Sales

Independent Variable	Totals	First Diff.	First Diff.	GMM FE, 3+ Inst.	GMM FE, 0+ Inst.	GMM (IV) FE, 0+ Inst.	IV FE, 0+ Inst.
1981-1985: 447 Firms							
Log L	.548(.026)	.630(.061)	.573(.064)	1.213(.177)	1.049(.108)	.995(.125)	1.095(.204)
Log C	.307(.018)	-.028(.036)	.152(.048)	-.124(.086)	-.081(.061)	-.044(.067)	-.087(.090)
Log K	.103(.009)	.018(.036)	.075(.035)	-.085(.080)	-.063(.041)	-.056(.045)	-.092(.062)
std err	.487	.129	.127				
χ^2(d.f.)				35.5 (27)	67.8 (63)	48.0 (45)	17.6 (9)
1985-1989: 381 Firms							
Log L	.543(.026)	.752(.050)	.700(.055)	.425(.118)	.467(.084)	.445(.096)	.126(.160)
Log C	.356(.018)	-.066(.039)	.063(.042)	.015(.112)	.039(.066)	-.010(.082)	-.183(.118)
Log K	.078(.011)	-.132(.052)	.034(.055)	-.162(.102)	-.138(.044)	-.105(.054)	-.193(.077)
std err	.475	.148	.149				
χ^2(d.f.)				27.5 (27)	72.1 (63)	51.3 (45)	5.7 (9)

The variables are the same as in table 2, with capital stocks at the beginning of year for the first two columns, and at the end of year for the five last ones. All equations contain year dummies and industry dummies at the level given in table 1.
Heteroscedastic-consistent estimates of the standard errors are shown in parentheses.
The method of estimation in the first 3 columns is ordinary least squares. In the last 4 columns two-step GMM is used. The column labelled IV uses only those moment conditions that are implied by the usual pooled IV estimator, while that labelled GMM(IV) uses the same moments, but imposes them for each year of data separately. See the text for a description of the instruments.

TABLE 5
Estimates for the Balanced US Panels, Dependent Variable: Log Sales

Independent Variable	Totals	First Diff.	First Diff.	GMM FE, 3+ Inst.	GMM FE, 1+ Inst.	GMM (IV) FE, 1+ Inst.	IV FE, 1+ Inst.
1981-1985: 535 Firms							
Log L	.576(.015)	.650(.028)	.555(.028)	.877(.130)	.897(.076)	.881(.093)	1.26 (.22)
Log C	.358(.012)	.088(.027)	.190(.029)	.218(.143)	.144(.090)	.126(.103)	.012(.156)
Log K	.035(.007)	.027(.032)	.120(.035)	.009(.102)	.033(.061)	.031(.067)	-.062(.083)
std err	.325	.129	.127				
χ^2(d.f.)				42.0 (27)	81.7 (51)	62.7 (33)	10.8 (6)
1985-1989: 442 Firms							
Log L	.639(.015)	.542(.030)	.423(.032)	.317(.116)	.384(.076)	.489(.104)	.552(.160)
Log C	.313(.015)	.138(.030)	.212(.035)	.336(.130)	.172(.085)	.099(.109)	-.057(.163)
Log K	.041(.008)	-.033(.040)	.147(.045)	-.139(.091)	-.039(.048)	.006(.059)	.069(.084)
std err	.343	.132	.130				
χ^2(d.f.)				32.8 (27)	61.2 (51)	49.7 (33)	21.1 (6)

See table 4 for precisions on the variables and methods of estimation.

Tables 4 and 5 also reveal an interesting fact about the efficiency gain from using GMM estimation for panel data rather than the conventional IV methods with stacked data. The column before last in the tables gives the GMM estimates (noted GMM (IV)) using as instruments only the lags 0 to 3 of the variables (for France) or the lags 1 to 3 (for the US), while the last column gives the standard IV estimates computed by using a GMM estimator only on the moment conditions corresponding to the stacked model.[18] All standard error estimates shown are robust, computed using the estimated variance of the orthogonality conditions from the model with lag 3+ instruments only, and so they can be directly compared. What is clear from the tables is that adding the longer lags as instruments where available produces little or no efficiency gain (because these additional lags are highly correlated with the instruments already present), but that switching from IV to GMM comes close to halving the standard errors. The reason is simple: GMM allows the projection on the instruments to be different for every year, whereas IV constrains it to be the same. In the traditional two stage least squares interpretation, GMM is using many more predictor variables, which implies a better predictor for the endogenous variables in finite samples, and hence smaller standard errors in general. Which estimator is preferred depends somewhat on what we are willing to assume about the process generating our right-hand-side variables. Because there is no reason to assume that it will be the same every year when we include longer and longer lags as instruments, it seems plausible to allow the projection coefficients to vary. We suspect that the case here is not atypical, and that the efficiency gain from GMM in panel data is coming primarily from the different year instruments and not simply from the additional lags.[19]

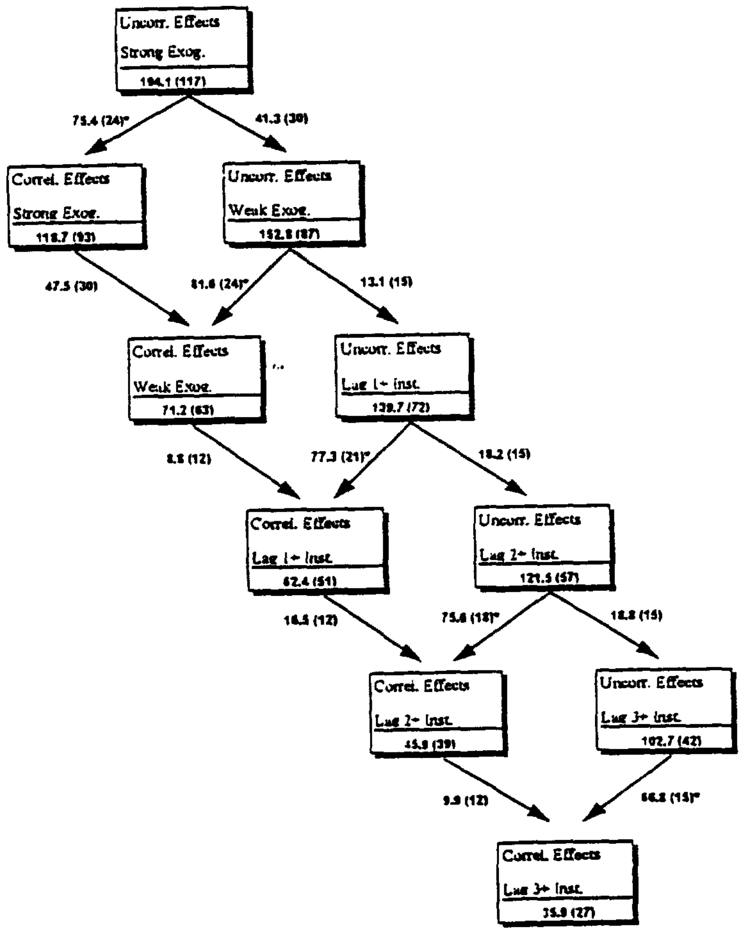
---

rates of labor, capital, and R&D capital in the first period, and .089, .221, and .543 in the second period. For the United States, the numbers are .031, .107, and .444 for the first period and .018, .058, and .451 for the second.

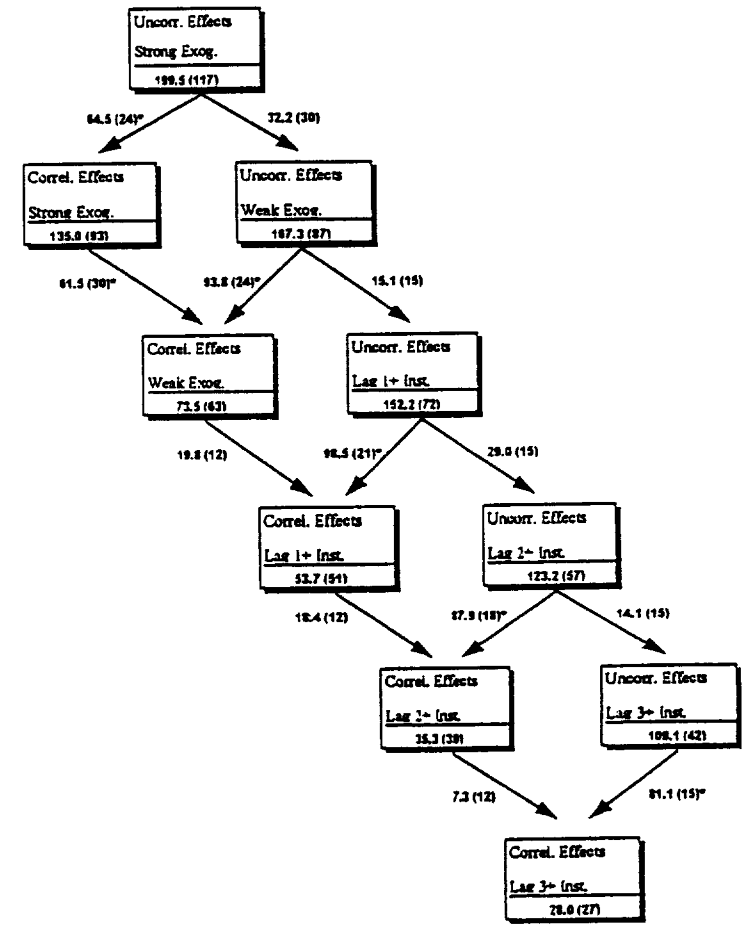
[18] It can be shown that the IV moment condition corresponding to the orthogonality conditions $\left[\sum_{i=1}^{N} (u_{it} \otimes z_i) = 0,\ t = 1,\ldots,T\right]$ is $\sum_{t=1}^{T} \sum_{i=1}^{N} (u_{it} \otimes z_i) = 0$. That is, the equivalent IV conditions just sum the relevant moment conditions over all the years in the panel. In the French case for example, there are 48 OCs in the second to the last column (= 4 years times 3 instruments times 4 lags). IV reduces this to 12 (=3 instruments times 4 lags).

[19] Note, however, that the small sample biases may become severe with an increasing number of instruments. See Bound, Jaeger and Baker (1993).

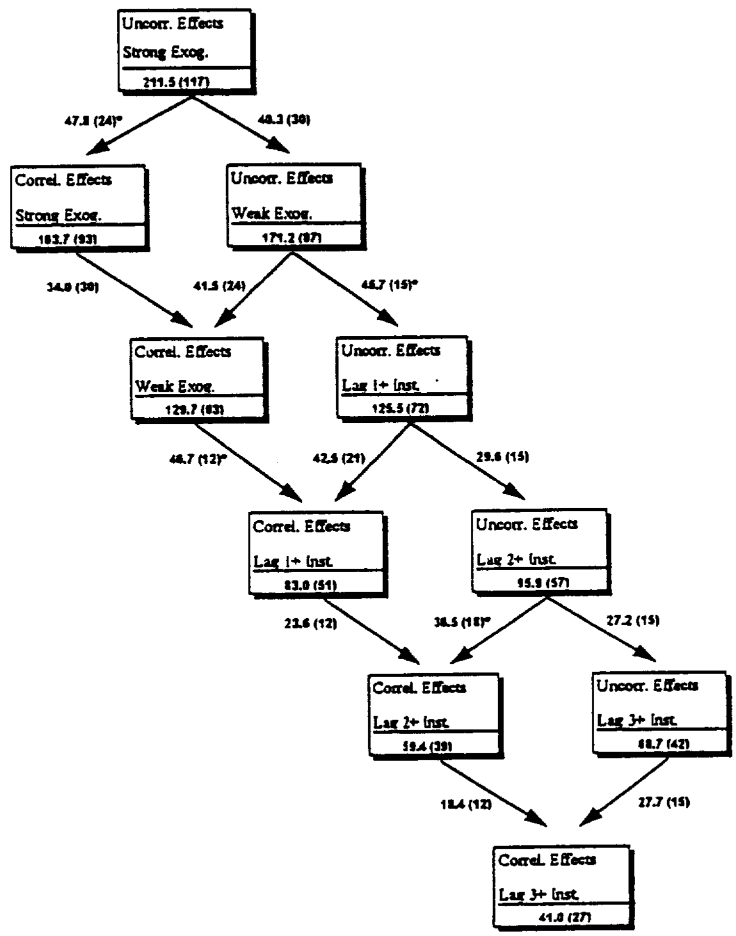
GRAPH 1
Testing for Exogeneity and Correlated Firm Effects
447 French Manufacturing Firms 1981-1985



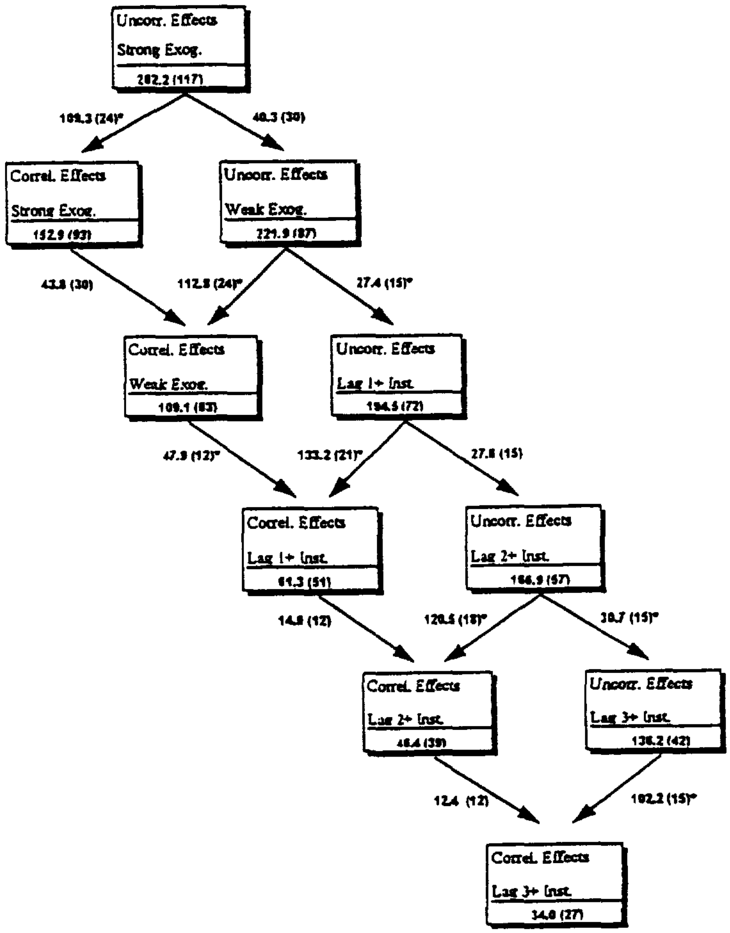
GRAPH 2
Testing for Exogeneity and Correlated Firm Effects
381 French Manufacturing Firms 1985-1989



GRAPH 3
Testing for Exogeneity and Correlated Firm Effects
535 U.S. Manufacturing Firms 1981-1985



GRAPH 4
Testing for Exogeneity and Correlated Firm Effects
442 U.S. Manufacturing Firms 1985-1989



5. Conclusions

In analysing the regressions presented in tables 2 through 5 and their many variants, we have reached the following conclusions: 1) The R&D contribution to sales or productivity growth during the 1980s seems to be somewhat lower in the 1980s than it was in the 1970s in both countries. 2) Using sales instead of value added does not seriously bias the results, although attempting instead to correct for materials inputs may. 3) The dating of the capital stocks makes a huge difference to their estimated coefficients. All of the estimates shown in table 2 are based on beginning of period stocks, and the end of period estimates would be about half again as high. This suggested simultaneity between both kinds of investment and sales (due to demand or liquidity shocks). Although this explanation seems to hold for the United States, where we reject the exogeneity of capital in the production function equation, it is less persuasive for France, where contemporaneous instruments are accepted. We advanced the tentative hypothesis that the difference in behaviour across countries is due to the fact that liquidity constraints impact US R&D-doing firms more than French firms. 4) Finally, the capital stock coefficient in the within firm estimates in France is zero (or negative), which we regard as a puzzle, and possibly a warning about the data.

With respect to the main substantive finding of this paper, we want to stress the two following points. First, the measured R&D elasticity using data which is not corrected for double counting ought to be zero if R&D investment is earning a "normal" rate of return, comparable to that of other inputs. It is not unlikely that this is the situation in the United States during the 1980s, when total industrial R&D was not growing that quickly. Hall (1993a) finds that the stock market valuation of the same firms' R&D capital has fallen by a factor of 4 during the 1980s, which implies that the market has perhaps recognized the relatively low private productivity of such investment.

Second, the results for the United States that use a sales measure deflated by a two-digit level deflator and adjust to some extent for quality change are suggestive: here we find a very substantial excess contribution of R&D to productivity growth (0.10 within firm). The implication is that R&D investment has been very productive in increasing "true" output, but that most of the gains have gone to the consumer in the form of lower prices. Because the price indices for the United States are only partially corrected for quality change, and because those for France have essentially no such correction, we regard this last finding as an invitation for future research that attempts to focus on the quality as well as quantity of output.

Appendix - GMM Estimation for a Linear Panel Data Model with Fixed Effects

A - General Methodology

The general methodology described here is based on that of Arellano and Bond (1991), Arellano (1988), and Schmidt, Ahn and Wynkowski (1992). The authors of these papers categorise the moment conditions implied by the use of linear regression models with predetermined rather than exogenous right hand side variables in a panel data setting and suggest estimating this type of model by means of the generalized method of moments (GMM). GMM is appealing in this setting because it is robust to heteroskedasticity across firms and correlation of the disturbances within firms over time, and can be efficient under fairly weak assumptions on the disturbances.

The most general model being considered is a random coefficients model with correlated effects:

$$y_{it} = x_{it}\beta_i + \alpha_i + \varepsilon_{it} \qquad i = 1,\ldots,N;\ t = 1,\ldots,T$$

$$y_{it} = x_{it}\bar{\beta} + v_i x_{it} + \alpha_i + \varepsilon_{it} \qquad \text{where } v_i = \beta_i - \bar{\beta}$$

The overall means of the data have been removed, so that we can assume $E[\alpha_i] = 0$ without loss of generality. The maintained stochastic assumptions are the following:

$$\begin{array}{ll} E[\varepsilon_{it}] = 0 & E[\varepsilon_{it}^2] = \sigma^2 \\ E[v_i] = 0 & E[v_i^2] = \sigma_v^2 \\ E[\varepsilon_{it}\varepsilon_{js}] = 0 & \text{for } i \neq j \text{ and for all } t,s \\ E[v_i v_j] = 0 & \text{for } i \neq j \\ E[\alpha_i \alpha_j] = 0 & \text{for } i \neq j \end{array}$$

However, we do not necessarily wish to assume that $E[\alpha_i | x_{i1}, \ldots, x_{iT}] = 0$. If we define a new composite disturbance as

$$u_{it} = v_i x_{it} + \alpha_i + \varepsilon_{it}$$

and its first difference as

$$\Delta u_{it} = v_i \Delta x_{it} + \Delta \varepsilon_{it}$$

then the covariance matrix for $\Delta u_i' = (\Delta u_{i1}, \Delta u_{i2}, \ldots, \Delta u_{iT})$ is

$$E[\Delta u_i \Delta u_i' | x_i] = \sigma_v^2 \Delta x_i \Delta x_i' + E[\Delta \varepsilon_i \Delta \varepsilon_i']$$

where Δx_i and $\Delta \varepsilon_i$ have been defined analogously to Δu_i. Even without the presence of the α_i, this covariance matrix is intrinsically heteroskedastic as well as non-diagonal, once we allow for random coefficients. Thus we prefer GMM estimation on the moments of u_{it} or Δu_{it} with lagged and current values of the x's; which lags are chosen is a subject for exploration in the body of the paper.

Our general procedure is to begin with a fairly weak assumption about orthogonality between the disturbance and the lagged right hand side variables, such as orthogonality between x's lagged three times and contemporaneous disturbances, and then to add more recent lags until the additional restrictions are rejected by a chi-squared test on the orthogonality conditions. To be more precise, let all our moment conditions be defined as

$$E[u_i(\beta) \otimes z_i] = 0$$

where

$$u_i(\beta) = [u_{i1}(\beta), u_{i2}(\beta), \ldots, u_{iT}(\beta)] \quad \text{and} \quad u_{it}(\beta) = y_{it} - x_{it}\beta$$

$$z_i = [z_{i1}, \ldots, z_{im}] \text{ and } m = \text{number of instruments per year}$$

The sample equivalents of these moment conditions are

$$f(\beta) = \frac{1}{N}\sum_{i=1}^{N} f_i(\beta) = \frac{1}{N}\sum_{i=1}^{N} u_i(\beta) \otimes z_i$$

The GMM estimator of β minimises the quantity

$$\phi(\beta) = f'(\beta)\, A\, f(\beta)$$

with respect to β, where A is a positive definite symmetric matrix. If A can be chosen as a consistent estimate of the inverse of the covariance matrix Ω of $f(\beta)$, this estimator will be consistent and asymptotically efficient. Even if A is inconsistent, the estimates of β will be consistent under fairly general conditions (Amemiya, 1977; Chamberlain, 1984; Hansen, 1982).

To estimate Ω and A consistently, we obtain estimates of β by a consistent method for our maintained model (the one using the smallest number of orthogonality conditions), and compute the estimated residuals based on these β's:

$$\hat{u} = y_{it} - x_{it}\hat{\beta} \qquad i = 1,\ldots,N;\ t = 1,\ldots,T$$

The sample covariance of $\hat{u}_i \otimes z_i$ is then computed; note that it is necessary in computing this covariance that the sample means of $\hat{u}_i \otimes z_i$ be removed, because we do not assume that $E[u_i \otimes z_i] = 0$ for all moments of u and z under the maintained model. Alternative methods for obtaining estimates of Ω and A are discussed below.

When we implement this GMM estimator for panel data, the z_i vector is the entire list of potential instruments $(x_{i1}, x_{i2}, \ldots, x_{iT})$ and the $u_i(\beta)$ are actually $\Delta u_i(\beta)$, the set of equations for the disturbances in the first differenced version of the maintained model. To test $E[\alpha_i \mid x_{i1}, \ldots, x_{iT}] = 0$, we add a set of moment conditions of the form $E[u_{it}(\beta) \otimes z_i] = 0$ to the set of differenced equations. This is equivalent to estimating the entire set of moment conditions in levels, but in this form the maintained model is clearly nested within the model which does not have correlated effects.

Until now, the Kronecker product notation, familiar from the work of Hansen (1982) and Hansen and Singleton (1982) has been used to simplify presentation. However, this notation is not generally suitable for panel data, because each time period t has a different number of instruments available with which to form orthogonality conditions. Our solution is to use a selection matrix on the original complete set of moment conditions, so that those which are not valid for a particular specification are not constrained to be zero. Define a kT by 1 vector

S of zeroes and ones that selects the appropriate moments from $\frac{1}{N}\sum_{i=1}^{N} u_i \otimes z_i$

Then the solution to the problem

$$\min_{\beta} \; f'(\beta)\; diag(S)\; \hat{A}\; diag(S)\; f(\beta)$$

is equivalent to the GMM estimator based only on the moment conditions which are valid under various exogeneity assumptions. Although these two estimators of β are equivalent (both are consistent), they do not necessarily coincide, since it is possible that the unused moment conditions will contribute covariance elements to the estimated covariance of $f(\beta)$, and these will appear in the computation of $\hat{A}$. One way around this particular problem is to use *diag(S)* in the computation of the covariance matrix itself. This is what is done here.

B - Actual Implementation

Our implementation of the different GMM estimators for our specific model is very similar to that of Arellano and Bond (1991) and Blundell and Bond (1994), although we have been using TSP rather than Gauss.

After differencing to remove the fixed effects, the model is the following:

$$\Delta u_{it} = \Delta y_{it} - \Delta x_{it}\beta$$

The x's are not necessarily strictly or even weakly exogenous. Firms index by $i = 1,\ldots,N$ and years by $t = 1,\ldots,T$, with T'-T years of presample data available. There are instruments available (including lagged x's) called z (m in each year, $t = T-T'+1,\ldots,0,1,\ldots,T$). Define the following row vectors:

$$\Delta u_i = (\Delta u_{i1}, \Delta u_{i2}, \ldots, \Delta u_{iT}) \text{ and } z_i = (z_i^{(1)}, z_i^{(2)}, \ldots, z_i^{(m)})$$

where

$$z_i^{(m)} = (z_{i,T-T'+1}^{(m)}, z_{i,T-T'+2}^{(m)}, \ldots, z_{i,0}^{(m)}, z_{i,1}^{(m)}, \ldots, z_{i,T}^{(m)}) \qquad \text{for all } m.$$

There are mT' elements in z_i and T elements in Δu_i. We begin by assuming that all the z's are valid instruments. Then the TmT' orthogonality conditions for the linear model specified above are the following:

$$f_i(\beta) = \Delta u_i \otimes z_i$$

The GMM estimator minimizes $\phi(\beta)$ with respect to β where

$$\phi(\beta) = \left[\frac{1}{N}\sum_{i=1}^{N} f_i(\beta)\right] A \left[\frac{1}{N}\sum_{i=1}^{N} f_i(\beta)\right]'$$

Asymptotically efficient estimates are obtained when $A = A_\infty$, the inverse of the true covariance of the moment conditions, or when A is replaced with a consistent estimator of A_∞. The GMM procedure in TSP uses the inverse of the sample covariance of the $f_i(\beta)$, evaluated at a consistent estimator of β (specifically, that from three stage least squares). Other consistent estimators of A can be computed if stronger assumptions are placed on the Δu's. We explore some of these possibilities here.

The sample covariance of the $\hat{f}_i = \Delta\hat{u}_i \otimes z_i$ is given by the following expression:[19]

$$A_N^{-1} = \frac{1}{N}\sum_{i=1}^{N}\left(\Delta\hat{u}_i\Delta\hat{u}_i' \otimes z_i z_i'\right)$$

[19] As noted above, because the time series structure of panel data means that the appropriate instruments are generally different for different time periods, in practice we will use a selection matrix to select the relevant moments from this Kronecker product.

The above equation suggests several ways to estimate the weighting matrix for GMM estimation:

(1) If the Δu_i are independently and identically distributed over time and across firms, then

$$plim\left[A_N^{-1}\right] = \sigma^2 I_T \otimes plim\left[\frac{Z'Z}{N}\right]$$

Therefore, use

$$A_N(1) = \left[\hat{\sigma}^2 I_T \otimes \frac{1}{N}\sum_{i=1}^{N} z_i z_i'\right]^{-1}$$

where $\hat{\sigma}^2$ is based on an initial consistent estimate of β (e.g, two or three stage least squares).

(2) If the Δu_i are serially correlated, but identically distributed across firms, then

$$plim\left[A_N^{-1}\right] = \Sigma \otimes plim\left[\frac{Z'Z}{N}\right] \qquad \text{where } \Sigma = plim\,[\Delta u\ \Delta u']$$

$$A_N(2) = \left[\hat{\Sigma} \otimes \frac{1}{N}\sum_{i=1}^{N} z_i z_i'\right]^{-1}$$

where $\hat{\Sigma}$ is again based on an initial consistent estimate of β.

(3) If the Δu_i are independent across firms, but serially correlated and heteroskedastic, use the full GMM weighting matrix (as in Hansen and Singleton, 1982):

$$A_N(3) = \left[\frac{1}{N}\sum_{i=1}^{N} \Delta\hat{u}_i \Delta\hat{u}_i' \otimes z_i z_i'\right]^{-1}$$

Estimates of β computed using these three estimates of A_N plus an estimate simply equal to the identity matrix are shown in table A.1. Estimates of the standard errors that are consistent even if A_N is not equal to the covariance of the orthogonality conditions are also shown in this table. Except for the identity matrix case, there is little difference between either the estimated coefficients or the standard errors across different choices of A_N for these data. This contrasts with findings reported in Blundell and Bond (1995) and may reflect our relatively large sample size.

TABLE A.1
Different Estimates of A_N, Balanced US Panel: 1986-1989

	Lag 3+ Instruments	Lag 2+ Instruments	Lag 1+ Instruments	Lag 0+ (Weak Exog.)	Strong Exogeneity
			$A_N = I$		
Log L	-.04 (.30)	0.27 (.10)	0.39 (.08)	0.46 (.07)	0.44 (.06)
Log C	0.78 (.38)	0.46 (.11)	0.32 (.09)	0.25 (.07)	0.39 (.06)
Log K	-.14 (.71)	-.09 (.08)	-.05 (.05)	-.04 (.05)	-.08 (.04)
Trace (DF)	35.4 (27)	51.6 (39)	67.2 (51)	103.2 (63)	166.9 (93)
P-value	0.128	0.085	0.064	0.001	0.000
Inst. Test (DF)	–	16.1 (12)	15.6 (12)	36.0 (12)**	63.8 (30)**
			$A_N(1) = \frac{\sigma^2}{N} [I_T \otimes Z' Z]$		
Log L	0.41 (.11)	0.30 (.10)	0.41 (.08)	0.46 (.07)	0.44 (.06)
Log C	0.53 (.15)	0.39 (.11)	0.29 (.08)	0.24 (.07)	0.33 (.06)
Log K	-.11 (.11)	-.06 (.08)	-.05 (.05)	-.04 (.05)	-.07 (.04)
Trace (DF)	36.2 (27)	51.2 (39)	66.8 (51)	103.5 (63)	166.8 (93)
P-value	0.11	0.09	0.068	0.001	0.000
Inst. Tes (DF)	–	15.0 (12)	15.6 (12)	36.8 (12)**	63.2 (30)**
			$A_N(2) = \Sigma \otimes \frac{1}{N} Z' Z$		
Log L	0.43 (.10)	0.32 (.10)	0.42 (.08)	0.47 (.07)	0.41 (.06)
Log C	0.42 (.12)	0.36 (.10)	0.28 (.08)	0.24 (.07)	0.33 (.06)
Log K	-.10 (.08)	-.05 (.08)	-.04 (.05)	-.04 (.05)	-.07 (.04)
Trace (DF)	39.6 (27)	51.1 (39)	66.6 (51)	103.7 (63)	166.7 (93)
P-value	0.056	0.092	0.070	0.001	0.000
Inst. Test (DF)	–	16.1 (12)	15.5 (12)	37.0 (12)**	63.1 (30)**
			$A_N(3) = \frac{1}{N} \sum_{i=1}^{N} [\Delta u_i \Delta u_i' \otimes z_i z_i']$		
Log L	0.33 (.10)	0.33 (.10)	0.42 (.08)	0.47 (.07)	0.44 (.06)
Log C	0.33 (.10)	0.34 (.10)	0.28 (.08)	0.24 (.07)	0.33 (.06)
Log K	0.00 (.08)	-.02 (.08)	-.04 (.05)	-.04 (.05)	-.07 (.04)
Trace (DF)	32.8 (27)	51.2 (39)	66.6 (51)	103.7 (63)	166.7 (93)
P-value	0.205	0.091	0.070	0.001	0.000
Inst. Tes (DF)	–	18.5 (12)	15.4 (12)	37.1 (12)**	63.0 (30)**

Notes: see next page

Notes:
All standard error estimates are robust to the presence of heteroskedasticity.
The row labelled "Inst. Test" is a chi-squared test for the validity of the additional instruments in the corresponding column, relative to the column on the left. This test statistic is distributed asymptotically as a chi-squared random variable with degrees of freedom equal to the number of additional moment restrictions under the null that all of the moment restrictions used hold. ** denotes values of the statistic for which the p-value is less then 0.01.

References

Amemiya, T. (1977), "The Maximum Likelihood and the Three Stage Least Squares Estimator in the Crest Nonlinear Simultaneous Equations Model", *Econometrica*, No. 45.

--- and T. MaCurdy (1986), "Instrumental Variable Estimation of an Error-Components Model", *Econometrica*, No. 54.

Arellano, M. (1988), "An Alternative Transformation for Fixed Effects Models with Predetermined Variables", Oxford University Applied Discussion Paper No. 57, Institute of Economics and Statistics, Oxford University.

--- and S. Bond (1991), "Some Tests of Specification for Panel Data: Monte Carlo Evidence and an Application to Employment Equations", *Review of Economic Studies*, No. 58.

Blundell, R.S., S. Bond, M. Devereux, and F. Schiantarelli (1992), "Investment and Tobin's Q", *Journal of Econometrics*, No. 51.

--- and S. Bond (1995), "Initial Conditions and Moment Restrictions in Dynamic Panel Data Models", University College (London)/Nuffield College (Oxford), mimeographed.

Bound, J., Z. Griliches, and B.H. Hall (1986), "Wages, Schooling, and IQ of Brothers and Sisters: Do the Family Factors Differ?", *International Economic Review*, No. 27.

---, D.A. Jaeger, and R. Baker (1993), "The Cure can be worse than the disease: A Cautionary Tale Regarding Instrumental Variables", National Bureau of Economic Research Technical Working Paper No. 137, Cambridge, Mass..

Chamberlain, G. (1984), "Panel Data", in Griliches, Z, and M. Intriligator (eds.), *The Handbook of Econometrics*, Volume II, Amsterdam, North Holland.

Cuneo, P. and J. Mairesse (1984), "Productivity and R&D at the Firm Level in French Manufacturing", in Griliches, Zvi, (ed.), *R&D, Patents, and Productivity*, Chicago, Ill., University of Chicago Press.

Gray, W. (1992), "TFP: The Productivity Database", Cambridge, Mass.: National Bureau of Economic Research, Diskette.

Griliches, Z. and J.A. Hausman (1986), "Errors in Variables in Panel Data", *Journal of Econometrics*, vol. 31.

--- and J. Mairesse (1983), "Comparing Productivity Growth: An Exploration of French and U.S. Industrial and Firm Data", *European Economic Review*, vol. 21.

--- and J. Mairesse (1990), "R&D and Productivity Growth: Comparing Japanese and U.S. Manufacturing Firms", in Hulten, Charles R. (ed.), *Productivity Growth in Japan and the United States*,, University of Chicago Press, Chicago, Illinois.

--- and J. Mairesse (1995), "Production Functions: The Search for Identification", National Bureau of Economic Research Working Paper No. 5067, Cambridge, Mass.

Hall, B.H. (1993a), "The Stock Market Valuation of Research and Development Investment During the 1980s", *American Economic Review*, vol. 83.

--- (1993b), "Industrial Research During the 1980s: Did the Rate of Return Fall?", *Brookings Papers on Economic Activity*. Micro Economics (2), Washington D.C.

--- (1991), "R&D Investment at the Firm Level: Does the Source of Financing Matter?", mimeographed.

--- (1990) "The Manufacturing Sector Master File: 1959-1987", National Bureau of Economic Research Working Paper No. 3366, Cambridge, Mass.

--- and J. Mairesse (1995), "Exploring the Relationship between R&D and Productivity in French Manufacturing Firms", *Journal of Econometrics*, vol. 65.

Hall, R.E. (1988) "The Relation between Price and Marginal Cost in U.S. Industry", *Journal of Political Economy*, vol. 96.

Hansen, L.P. (1982) "Large Sample Properties of Methods of Moments Estimators", *Econometrica*, vol. 50.

--- and K. Singleton (1982), "Generalized Instrumental Variables Estimation of Nonlinear Expectations Models, *Econometrica*, vol. 50.

Holtz-Eakin, D., W. Newey, and H. Rosen (1988), "Estimating Vector Autoregressions with Panel Data", *Econometrica*, vol. 56.

Keane, M.P. and D.E. Runkle (1992), "On the Estimation of Panel Data Models with Serial Correlation When Instruments are not Strictly Exogenous", *Journal of Business and Economic Statistics*, vol. 10.

Mairesse, J. and M. Sassenou (1991), "R&D and Productivity: A Survey of Econometric Studies at the Firm Level", *Science-Technology-Industry Review*, No. 8. Paris, OECD.

OECD (1991), *Basic Science and Technology Statistics*, OECD, Paris.

Schankerman, M. (1981), "The Effect of Double Counting and Expensing on the Measured Returns to R&D", *Review of Economics and Statistics*, vol. 63.

Schmidt, P., S.C. Ahn, and D. Wyhowski (1992), "Comment on Keane and Runkle", *Journal of Business and Economic Statistics*, vol. 10.