

Analysing Firm-Level Labour Productivity Using Survey Data*

Mark Rogers and Yi-Ping Tseng
Melbourne Institute of Applied Economic and Social Research
The University of Melbourne

Melbourne Institute Working Paper No. 10/00

ISSN 1328-4991

ISBN 0 7340 1490 2

October 2000

* This paper is the result of work being undertaken as part of a collaborative research program entitled 'The Performance of Australian Enterprises: Innovation, Productivity and Profitability'. The project is generously supported by the Australia Research Council and the following collaborative partners: Australian Tax Office, Commonwealth Office of Small Business, IBIS Business Information Pty Ltd, Productivity Commission, and Victorian Department of State Development. The views expressed in this paper represent those of the authors and not necessarily the views of the collaborative partners.

Melbourne Institute of Applied Economic and Social Research
The University of Melbourne

Victoria 3010 Australia

***Telephone* (03) 8344 5330**

***Fax* (03) 8344 5630**

***Email* melb.inst@iaesr.unimelb.edu.au**

***WWW Address* <http://www.ecom.unimelb.edu.au/iaesrwww/home.html>**

Abstract

This paper investigates the determinants of firm-level labour productivity in the manufacturing sector using GAPS data. These data are from a stratified survey, where the strata are based on industry and firm size. The paper focuses on whether weights should be applied in the regression analysis. Augmented Cobb-Douglas production functions are estimated, where a set of dummies are used as proxies for firm-level knowledge stocks. The regression results show that there are significant differences between the parameters estimated by weighted least squares (WLS) and OLS, particularly for the variables union density and training expenditure. These differences can be caused by parameter heterogeneity (across strata); in theoretical terms this means that applying the same production function across all firms is not appropriate. Given this parameter heterogeneity, both the OLS and WLS methods do not estimate parameters of interest. Instead, there is a requirement to estimate sub-sample regressions. These are presented in the second part of the empirical results.

Key words: *labour productivity, weights, survey*

1. Introduction

This paper uses data from the ABS Growth and Performance Survey to analyse the determinants of labour productivity in manufacturing firms. Two specific issues are investigated. First, how should survey data be used to estimate economic models of firm productivity? In particular, given that the sampling method to create the survey was based on strata (industry and firm size), should a weighted least squares or an ordinary least squares estimator be used? Second, how can we use the available data to test the idea that the stock of knowledge capital varies across firms? The firm's stock of knowledge is related to past investments in R&D, innovation, human capital and organisational capabilities. At a general level, these factors are often thought to play an important role in a firm's level of productivity but it is often difficult to measure their influence, hence they are rarely included in productivity analyses. The GAPS survey includes some questions that can be used to assess such issues.

The structure of the paper is as follows. The next section outlines the theoretical framework used in the paper. This is an augmented Cobb-Douglas production function approach which is common in the literature. This framework suggests that the most important determinant of labour productivity will be the capital to labour ratio. Section 3 provides a brief overview of some related empirical studies on firm-level productivity. This covers the set of additional variables that augment the Cobb-Douglas production function. These relate to firm investment in innovation, R&D, training, computers, as well as whether the firm is unionised, foreign owned, or exports some of its output. Section 4 provides a discussion of the data and variables used. Section 5 provides a summary of the issues of weighting in regression analysis (a more detailed discussion is confined to an Appendix). Section 6 contains the regression results, and section 7 concludes.

2. Theoretical framework

The most common method of empirically analysing productivity is to start from a production function. A production function, in its most general form, links output (Y) to the range of inputs (X), where both can be vectors. For empirical analysis, a specific functional form of the production function needs to be specified. Following Griliches (1986), assume that a firm's value added is determined by the (Cobb-Douglas) production function

$$V = AK^\lambda C^\alpha L^\beta, \quad [1]$$

where V is value added¹, C is physical capital, L is labour, K is level of knowledge capital, and A is a constant. The unusual aspect of [1] is the presence of ‘knowledge capital’. In many studies, including Griliches (1986), this is taken to be the ‘accumulated and still productive’ research capital as derived from previous R&D expenditures. Using this approach empirical studies have tried to estimate the private and social returns to R&D investment. In this paper, ‘knowledge’ is interpreted more broadly to include past investments in innovation, organisational techniques, human capital of both managers and workers (where human capital refers to the accumulated education, training and experience), in addition to R&D investment. All of these types of investment can potentially influence the value added of a firm. Ideally, an empirical study should have separate, time series data on each of these elements of the firm’s knowledge capital. However, in general these data are not available, often leading researchers to focus on a limited form of [1], perhaps with just C and L present.

An alternative to excluding knowledge capital altogether is to use any available data to proxy which firms have higher levels of such capital. In this study, data from the GAPS study are used to create a number of dummy variables to proxy high and low knowledge firms. Specifically, taking natural logs of [1], and assuming that $\alpha + \beta = 1$ (i.e. constant returns to physical capital and labour) we have

$$\ln \frac{V}{L} = \ln A + \lambda \ln K + \alpha \ln \left[\frac{C}{L} \right] = \ln A + \alpha \ln \left[\frac{C}{L} \right] + \beta_i D_i \quad [2]$$

where D_i represents dummy variables that are intended to capture whether a firm has a high or low knowledge stock. There are data on R&D expenditures for two years prior to productivity data. However, this is insufficient to calculate a R&D capital stock, although the regressions do include a lagged R&D intensity measure for a single year as a partial test of the role of R&D.

¹ Equation [1] is specified in terms of value added rather than output which implies materials do not enter the production function in the same way as labour or capital.

The limited ability of the data to measure the true nature of the production function means that the empirical analysis should be viewed as exploratory. Moreover, it also implies that assuming a single production function applies to all firms is unwise. Theoretically, the use of proxy variables for the knowledge stock implies that the estimated production function should, as far as possible, be applied to a set of homogenous firms, for example, those in a single industry. (i.e. a dummy variable distinguishing between ‘high’ and ‘low’ knowledge firms may well have a different coefficient in different industries). Thus, the approach used here is summarised below:

- A simple Cobb-Douglas production function is used to represent the production technology of the firm. This approach suggests that regressions should only include ‘homogenous’ firms.
- The production function implies a major determinant of labour productivity is the capital to labour ratio.
- Labour productivity will vary across firms, with the same capital to labour ratio, to the extent that the knowledge stock varies. Although data constraints mean that the knowledge stock cannot be accurately measured, a series of proxies can be used for the knowledge stock. Under the (unrealistic) assumption of costless and instantaneous knowledge diffusion, and that the proxies are appropriate, none of these proxies should be significant.

3. A review of empirical productivity analyses

There is extensive literature on the influence of R&D on productivity. As indicated above, the normal procedure is to construct a R&D capital stock by using historical data on R&D (and an assumed rate of depreciation). At the firm-level, most studies have followed the Griliches (1986) approach detailed above. In general, studies find that R&D has a positive and significant effect on productivity (for reviews see Cohen, 1995, Griliches, 1995, Nadiri, 1993, and Mohnen, 1992), with estimated private rates of return often being between 20% and 30%. These studies normally have access to time series data on the R&D expenditures of firms and can construct an R&D capital stock (by assuming a rate of depreciation or obsolescence – often 10% or 15%). Use of R&D capital stock data is preferred since the production relationship is specified in stocks. There

is also an implicit lag time between R&D effort and its benefits. Studies suggest that the lag structure is bell shaped with a mean lag time of between 4 to 6 years. This situation means, ideally, that a long time series on firm-level R&D should be available, something that is rarely available in practice. GAPS data only provides 3 years of data on R&D implying that a full investigation of role of R&D on productivity is not possible. Thus only the lagged innovation status is included in this study as an indicator of R&D results.

One issue in recent studies of firm-level productivity is the role of information technology (IT) in determining productivity. The rapid rise in expenditures on IT contrasts with relatively little evidence of its productive impact, leading some commentators to suggest a 'productivity paradox', with the implication being that firms are sub-optimally investing in IT.² However, some recent firm-level analysis suggests that IT investment raises productivity. Brynjolfsson and Hitt (1995) use data on around 200 Fortune 500 US firms for the period 1988 to 1992, finding that a composite 'IT capital and labour' stock measure is positively and significantly related to value added. This study is of additional interest since they found support for using the (simpler) Cobb-Douglas approach, rather than a translog specification, and also that coefficients do vary across industries. Lichtenberg (1993) provides further empirical evidence of the importance of IT capital using similar data. The GAPS data does not allow us to construct an IT capital stock measure so we cannot closely follow previous IT studies. Instead, dummy variables are constructed for how long the firm has used computers and also the ratio of employees using computers to all employees.

There is a substantial empirical literature on the role of unions and labour productivity. In short, some arguments would suggest that unions may increase restrictive work practices and industrial disputes thereby lowering productivity (this is called the "monopoly face" of unions). In contrast, there are arguments that unions may raise communications within the firm and also reduce labour turnover (the "collective voice" view). Reduced labour turnover may lead to more experienced and better trained workers, hence higher productivity. Improved communications may also reduce inefficient practices and, possibly, boost innovative activity within the firm, again raising productivity. Given the two opposing theoretical views, the empirical issue is

² Baily and Gordon (1988) provide an early discussion of these issues; Triplett (1999) summarises the latest views of this 'paradox'.

which one dominates. Empirical studies have not found a consistent pattern. In Australia, both Crockett et al (1992) and Drago and Wooden (1992) find some evidence of a negative effect of unions on productivity using the AWIRS data.

Another possible factor in explaining productivity is the export status of the firm. Firms that export may be under more competitive pressure, hence they may eliminate inefficiencies and increase productivity. This type of argument cannot be pressed to far since some firms producing for the domestic market may also be under substantial pressure from imports (and, implicitly, international firms); equally some export markets may not require high efficiency, especially if the exports are derived from natural resources found only in the domestic country. Aw and Hwang (1995) use data for 2382 Taiwanese electronics firms to assess the differences between exporters and non-exporters. They test whether the two groups of firms can be pooled to estimate a production function, finding that they cannot. Moreover, they find that exporting firms tend to have higher productivity than non-exporters (although this result does not hold for all product sectors). This result they attribute to both the ‘competitive’ conditions argument above and also that exporters can benefit from “the transmission or diffusion of new or improved technology to exporters from foreign buyers [of these products]” (p.328).

The important role of human capital and training in productivity growth is widely recognised at the economy level (see Barro and Sala-i-Martin, 1995). Various cases studies also suggest an important role for human capital and training factors (see, for example, Mason and Finegold, 1997). Firm level evidence is less common, reflecting the lack of data sets that combine both firm level productivity measures and human capital or training data. A recent exception is Black and Lynch (1996) who use an augmented Cobb-Douglas production function to analyse various aspects of human capital and training. They use survey data of around 3000 US firms in 1993. They find productivity is higher in firms that have a higher average employee education level. With respect to training they find a mixed story. In manufacturing, formal training outside of work hours has a positive association with productivity, while other measures of training – overall effort, computer training, teamwork training, supervisor training – have no significant associations. For non-manufacturing firms they only find a role for computer related training. They find unionised, non-manufacturing firms are more productive. They also use dummy variables for recruitment priorities, use of ‘total quality management’ (TQM), use of

benchmarking and whether the firm is an R&D centre. Note that this use of dummy variables to proxy more complex variables is one that we follow here.

4. Data

The data employed in this paper are from a balanced panel of the Australia Bureau of Statistics' Growth and Performance Survey (GAPS, also called the Business Longitudinal Survey, BLS) for the period 1994/5 to 1996/7.

The GAPS is constructed under the objective of examining the relationship between the characteristics and behaviour of firms and their performance over time. It includes a set of core questions, which were asked each year and a set of one-off questions addressing different policy issues each year. The core questions include employment, ownership, union membership, export status, business practice, financial structure and important information in the balance sheet, etc. The main specific topics addressed in the three surveys are innovation and training in 1994/5, labour turnover and business links in 1995/6 and the use of computers in 1996/7. This allows researchers to investigate different policy issues addressed in each survey using a single cross-section or to link all the information by taking advantage of the panel nature of the survey.

The survey design of GAPS is different from a normal panel survey. In the first year (1994/5), about 8700 businesses are selected from the ABS business register³ based on the stratified random sampling method, where the stratification was by both industry and employment size classification. In the subsequent years, the sample size has dropped to around 5600.

In the 1995/6 survey, the sample used in the 1994/5 survey was re-stratified into 2 categories: 'high' and 'normal' performance businesses. The 'high' performance category was based on whether the firm was innovative, exporting, or recorded increasing sales or employment over the period 1993 to 1995. All other firms are allocated to the 'normal' performance category. To assist analysis of the behaviour of high performance firms, all businesses in the first category are traced in this survey, which provided about 3,400 observations. About 2,200 businesses were randomly selected from the 5,600 businesses in the second category. In addition, a sample of

³ Businesses in public sector, agriculture, health and education, communication services and a number of smaller industries were not included in this survey.

about 800 “new” businesses was also selected from those had been added to the ABS business register since the previous GAPS collection. In total, around 5700 businesses are surveyed in the second-year GAPS survey.

In the 1996/7 survey, all the firms surveyed in the previous year were traced, although some businesses closed, of course, and dropped out of the panel. New businesses were also introduced to keep the sample representative in the scope of this survey.

The complex survey design of GAPS raises our interest in investigating the importance of applying sample weights in the regression analysis on the relationship between firm characteristics and firm performance. This will be discussed in detail in the next section.

The three years of survey data are used to form a balanced panel (which has about 4,500 businesses). Because the nature of the three sectors, manufacturing, non-manufacturing, and financial, are very different, this paper focuses on manufacturing firms to minimise the problem caused by the large differences of production structure. Although this has reduced the sample size to 1683, it still gives us sufficient observations to perform regression analysis.

The empirical model uses labour productivity in 1996/7 as the dependent variable and explanatory variables from the previous three years of survey data. The balanced panel data set, therefore, allows us to take advantage of information collected from some of the one-off questions in the 1994/5 or 1995/6 survey. In addition, the use of explanatory variables from prior to the period when the labour productivity is taken allows some degree of exogeneity. This approach, however, means we do not take advantage of panel data regression techniques, such as fixed or random effect models. Obviously, such models would not allow all the explanatory variables used below to be included, but they would offer a method of controlling for (unobserved) time invariant, firm specific factors that may influence productivity. The use of panel data models is left for future analysis. The description and summary statistics of important variables are presented in Table 1 and Table 2.

Table 1 **Summary Statistics**

Variable	Obs	Mean	Std. Dev.	Min	Max
Lnlp97	1683	4.18	0.66	-1.10	6.06
Lnkl97	1683	4.77	1.13	-0.83	9.13
Innov95	1683	0.41	0.49	0	1
Union2 (1-50%)	1683	0.25	0.43	0	1
Union3 (>50%)	1683	0.13	0.34	0	1
Train95	1683	0.40	0.80	0	10
Export97	1683	0.39	0.49	0	1
Foreign	1683	0.10	0.30	0	1
Ratio computer	1683	0.30	0.29	0	1

Table 2 **Description of variables**

Variable	Description
lnlp97	Log of labour productivity in 1997. Labour productivity is defined as value added per effective (i.e. full-time equivalent) employee in 1997. Numbers of part-time employees are recalculated into equivalent number of full-time employees based on the average working hours of part-time and full-time employees published by ABS.
Lnkl97	Log of labour capital ratio in 1997. Capital here is adjusted by the value of leasing capital and the operation hours per week.
Innov95	Innovation status in 1995. Businesses were identified as innovative if they developed or introduced any new or substantially changed products or processes.
Union2 (1-50%)	Union density dummies (up to 50% of employees are union members) in 1997. The base group of this set of union density dummies is those businesses with no employees who were union members.
Union3 (>50%)	Union density dummies (more than 50% of employees are union members) in 1997.
Train95	Training expenditure per effective employee in 1995.
Export97	Export status (yes = 1) in 1997
Foreign	Foreign ownership (yes=1) in 1997. Businesses are defined as foreign owned if more than 50% of the company equity is foreign owned.
Ratio computer	Ratio of the number of employees who use a computer at least once a week to total number of employees in 1997.

5. Estimating regression models from survey data

Equation [3] shows the basic linear regression model we attempt to estimate, where y is a vector of labour productivity, and ε is a vector of random error with mean 0 and variance σ^2 . X denotes a matrix of explanatory variables, including vectors of capital labour ratio, innovation, export status, foreign ownership, training expenditure, computer usage and union density dummies.

$$y = X\beta + \varepsilon \quad [3]$$

If a sample employed in regressions was a simple random sample or a stratified sample with equal probability of inclusion in the sample for all observations, it makes no significant differences whether we use weighted or unweighted data. Given that the GAPS data is a stratified sample with large variation in selection probability across observations whether weights should be implemented in the estimation becomes an issue.

In general, the decision to use weights or not depends upon the purpose of study, the parameters of interest and whether the heterogeneity of coefficients among different strata exists. If the regressions are to explore association by looking at the mean of one variable conditional on others, in general, weighted least square with corrected standard errors is preferred. If not, several different strategies can be applied to deal with heterogeneity and sampling design.

To be more specific, if we believe that β is homogeneous across strata, and the mean and variance of ε conditional on X and J (strata) are independent of (X, J) , both β_{OLS} and β_{WLS} are unbiased and consistent estimators of β . However, β_{OLS} is the most efficient estimators among all linear unbiased estimators and it therefore is preferred.

How do we test whether the model satisfies the above assumption or not? The standard approach is to test whether OLS estimates are significantly different from WLS estimates as suggested by DuMouchel and Duncan (1983). If the null hypothesis that $\hat{\beta}_{OLS} = \hat{\beta}_{WLS}$ fails, it implies β is not homogeneous across strata or there are omitted explanatory variables in the regression model. If we believe that the correlation between ε and (X, J) is caused by omitted variables, the solution is to search for extra predictors or to incorporate weights in the regression. However, β_{WLS} is a

consistent estimator of β^* which is the joint effect of the existing explanatory variables and the unobserved variables, but not the true β we intend to estimate.

If we believe there is heterogeneity of β across strata, the stratum-specific vector of parameters (β_s) can be obtained by sub-sample regressions.

If the parameter of interest is the population-weighted parameters: $\beta_p (= \sum_{s=1}^S \frac{N_s}{N} \beta_s)$, Deaton (1997) shows that neither β_{OLS} nor β_{WLS} is a consistent estimator of β_p (see appendix). As Deaton stresses, this is not due to a problem with the estimators; it simply reflects the heterogeneity in the population.

In the study of productivity, of course, it might be argued that there is little point in obtaining a consistent estimate of β_p since this might be of little use for policy work. For example, the average effect of training across the population might be found to be zero, even though for some industries it might be positive, while for others negative. In terms of policy making, it is more important to identify which industry has greater returns on training. Thus, if there is concern over parameter heterogeneity, the suggestion is to estimate separate regressions for each sub-sample (strata) and calculate β_p if necessary.

Nevertheless, problems may occur when running regressions by strata if the sample size in each stratum is not large enough. In this case, there is no alternative to combining some strata which are, theoretically, more likely to have homogeneous β .

6. Regression results

The previous section discussed the potential differences between OLS and WLS and when the weights should be used in the regression. In this section, first of all, we present both OLS and WLS regression results to give a hint on how different they are. Then we adopt the method suggested by DuMouchel and Duncan (1983) to test the statistical significant difference between the coefficients estimated by OLS and WLS. Then we choose the best estimation strategy based on the first two results to analyse labour productivity.

Table 3 below shows both the OLS and WLS for the full sample of manufacturing firms in the data. The results suggest there are important differences between the estimators. The coefficient on the capital to labour ratio is significant in both regressions with a value close to 0.4. However, the WLS results suggest that unions have a strong positive association with labour productivity, whereas the OLS shows no association. The OLS results suggest that training, foreign ownership and the ratio of employees using computers to all employees are also important factors, in contrast to the WLS results.

A test of the equality of the two sets of coefficients (i.e. $(\beta_{WLS} - \beta_{OLS})=0$) can be based on the “auxiliary” regression (see Deaton, 1997, p.72)

$$y = X\beta + WX\rho + \varepsilon \quad [4]$$

and using an F-statistic to test $\rho=0$.

Undertaking this test shows the differences are significant (F-stat = 2.09). In view of this, the econometric arguments suggest that either a) the existence of unobserved variables that are correlated with strata, or b) parameter heterogeneity may be a problem. In light of the theoretical issues discussed in section 2, our view is that parameter heterogeneity is the appropriate course to investigate. In other words, there appears little theoretical support for the assumption that the Cobb-Douglas production function should hold across all firms in the sample.

To investigate the issue of coefficient heterogeneity, a series of regressions are run on each two-digit industry. Theoretically, we should run separate regression by industry-firm size groups, that is individual strata, according to the survey design. Since the sample sizes are small in some strata, we have no choice but to combine some strata. Assuming that the parameters are more likely to differ across industries, regression sub-samples by industry are taken (although firm-size dummies are entered in the regressions to allow level effects in productivity across firm size).

Table 3 OLS vs WLS

Dependent variable: log of labour productivity

	OLS		WLS	
	coefficients	t-statistics	Coefficients	t-statistics
Lnk197	0.366	21.13	0.407	8.92
Innov95	-0.048	-1.93	-0.072	-1.40
Union2 (1-50%)	0.017	0.60	0.160	2.54
Union3 (>50%)	0.016	0.47	0.190	2.60
Train95	0.030	1.92	0.001	0.05
Export97	0.042	1.64	0.070	1.44
Foreign	0.086	2.17	0.068	1.06
Ratio computer	0.127	2.25	0.221	2.60
Industry dummies	Yes		Yes	
Firm size dummies	Yes		Yes	
Constant	2.215	23.63	2.012	9.26
Number of observations	1683		1683	
R-square	0.505		0.523	

Note: Industry dummies are defined at the 2-digit level.

The null hypothesis of equal coefficients across all industries was rejected (F-stat of 1.97 with degree of freedom (96, 1566)). The results for each industry are shown in Table 4 and indicate that there is some heterogeneity across industries. In particular, a general result is that none of the explanatory variables, apart from the capital to labour ratio, have a significant association with labour productivity across all industries. A summary of the results can be made as follows:

- The coefficient on the capital to labour ratio is always positive and significant with a magnitude of between 0.29 and 0.47.
- Innovation in 1995 rarely shows a significant association with labour productivity in 1997. The exceptions are for industry 23 (a positive association) and industry 27 (a negative association).
- The role of unions is almost never significant (the exception is industry 29, where union membership of between 1-50% has a positive association).

- The coefficient on the training intensity variable is positive and significant in industry 22, 25, and 26, and negative and significant in industry 29.
- Export status shows little association with productivity apart from a positive association in industry 24
- Foreign ownership status shows a positive and significant coefficient in industry 25 only.
- The ratio of computer users to non-users is only positive and significant in industry 24.
- The results for the firm size dummies vary substantially across industries. In industries 21, 23, 25 and 27 none of the firm size dummies have significant coefficients. In contrast, in industries 22, 26 28 and 29 the majority of the firm size dummies have significant coefficients.

The variation in the coefficient estimates across industries is interesting and worthy of further discussion. Consider first the variation in the coefficient on the capital to labour ratio. The inclusion of this variable is based on the Cobb-Douglas production function, hence the variation in the capital-labour coefficient can be taken as evidence that the elasticity of capital with respect to output does vary across industries. However, an alternative explanation is that the functional form used is not appropriate, with perhaps a translog or CES functional form being preferred (i.e. if a more flexible functional form for the production function was used parameters may not vary across industry). An additional observation is that the role of the firm size dummies may contain evidence on this issue. If the capital to labour ratio and firm size are positively correlated, then the firm size dummies may 'pick up' some of the capital to labour effect, especially in cases where the functional form is not appropriate.

A further result of interest is the general lack of significance for the coefficient on lagged innovation. If innovative activity either raises price, or improves the efficiency of the firm, we should expect a positive effect on labour productivity. A number of reasons may explain the lack of this result. First, the measure of innovation is a simple dummy variable and cannot control for quality (i.e. the value of the innovation) or quantity of innovation. Second, although we might, in

general, expect innovation to positively influence firm performance, this influence may operate through raising revenue or market share, rather than in raising value added per employee. Third, the lag times for the impact on labour productivity may be longer than allowed for here. Even given these reasons, the results are puzzling: in only one industry (23) is there an apparent positive association.

The results on the dummy variables for union density show that only one coefficient is significant. This is especially interesting since use of the WLS regressions showed a strong positive association between unions and labour productivity. This is an example of how different the results can be when using weights in the analysis. A task for further analysis is to investigate why the weighted results can be so different.

Table 4 OLS regressions by industry

Dependent variable: log of labour productivity

	Food, Beverage and Tobacco Manufacturing (21)	Textile, Clothing, Footwear, and Leather (22)	Wood and Paper Product Manufacturing (23)	Printing, Publishing and Recorded Media (24)	Petroleum, Coal, Chemical and Associated Product (25)	Non-Metallic Mineral Product Manufacturing (26)	Metal Product Manufacturing (27)	Machinery and Equipment Manufacturing (28)	Other Manufacturing (29)
Lnk197	0.4663**	0.3635**	0.4171**	0.3980**	0.3025**	0.2900**	0.3715**	0.3631**	0.4121**
Innov95	-0.0946	-0.0043	0.2292**	-0.0282	0.0039	-0.0589	-0.1556**	-0.0604	0.0405
Union2 (1-50%)	0.0567	0.0733	0.0528	-0.1060	0.0032	-0.1235	0.0429	0.0329	0.2421**
Union3 (>50%)	0.1221	-0.1013	0.0583	-0.1093	0.0915	-0.0084	0.0270	-0.0389	0.1681
Train95	-0.0111	0.0937**	0.0952	0.0494	0.1264**	0.1776*	-0.0081	0.0217	-0.1529**
Export97	0.0837	-0.0757	-0.1736	0.2055**	0.0612	-0.1724	0.0508	0.0340	0.0320
Foreign	0.1631	-0.0523	-0.0444	0.2204	0.2157**	-0.0658	-0.0473	0.0612	0.0845
Ratio computer size (5-19)	0.0268	0.4730	0.2296	0.2846**	-0.0604	0.3314	-0.0257	0.0666	0.1581
size (20-99)	0.2343	0.4600**	0.1294	0.0301	0.2064	0.4271*	0.1533	0.1897*	0.2428**
size (100-199)	0.1698	0.3288*	0.0951	-0.0790	0.2064	0.5044*	0.1429	0.2492**	0.2592**
size (200+)	0.2264	0.1613	0.0839	-0.3406*	0.0810	0.8409**	0.1282	0.3159**	0.2793*
Constant	0.1826	0.4437*	-0.0278	-0.2549	0.2570	0.8307**	0.2971	0.2000	0.1341
	1.6480**	1.9942**	1.9343**	2.2797**	2.4845**	2.3125**	2.3387**	2.2424**	1.9164**
Number of obs	199	144	99	138	219	90	217	416	161
R-squared	0.5266	0.5443	0.5860	0.6524	0.5743	0.5854	0.5395	0.4040	0.4699

Note: 1. “*” indicates significant at 10% level and “**” indicates significant at 5% level.

7. Conclusions

This paper has analysed the determinants of labour productivity in a sample of Australian manufacturing firms in 1997. The major aims of the paper are to highlight the econometric issues in using survey data, and also to investigate how survey data can be used to assess the importance of knowledge capital in productivity differences.

When data from a stratified survey is used for regression analysis there is an on-going debate over whether to use a weighted least square (WLS) or an ordinary least square (OLS) estimator. However, if the researcher expects that there is coefficient heterogeneity in the population neither of these approaches is valid. In the current case of investigating firm labour productivity there are strong theoretical reasons for expecting coefficient heterogeneity, since it appears unlikely that firms in different industries have the same productivity determinants. In view of this, the paper initially tested the equality of coefficients between the OLS and WLS estimators. Rejection of this test suggested that coefficient heterogeneity may indeed be an issue. To investigate this issue further sub-sample regressions, by industry, were run.

The industry regressions showed that there was considerable heterogeneity of coefficient estimates within the population. The only variable that was consistently significant and positive across all sub-sample regressions was the capital to labour ratio. Although the typical magnitude of the coefficient on the capital to labour ratio was, as expected, close to 0.4, the sub-sample regressions showed values between 0.47 and 0.29. There was little consistency in other estimates across sub-samples. For example, although the WLS results suggested higher unionisation was associated with higher labour productivity, this result was only supported in one industry (29).

In practice, it is difficult to clearly distinguish where the heterogeneity of β comes from. For example, our results show that the productivity increase due to training differs across industries, with only 3 out of 9 industries showing a positive and significant effect. An explanation of this heterogeneity may be that the average education level of employees in some industries, such as high technology industries, is higher than that in the others. The returns to training may be positive correlated with education level. Since the data does not contain employees' personal characteristics, we are unable to control for average education level of employees.

References

- Aw, B. and Hwang, A. (1995). "Productivity and the Export Market: A Firm-level Analysis". *Journal of Development Economics*, 47, 315-332.
- Baily, M. and Gordon, R. (1988). "The Productivity Slowdown , Measurement Issues, and the Explosion of Computer Power". *Brookings Papers on Economic Activity*, 2, 347-431.
- Barro, R. and X. Sala-i-Martin (1995). *Economic Growth*. New York, McGraw-Hill.
- Black, S. and Lynch, L. (1996). "Human-Capital Investments and Productivity". *American Economic Review*, 86(2), 263-267.
- Brynjolfsson, E. and L. Hitt (1995). "Information Technology as a Factor of Production: The Role of Differences Among Firms". *Economics, Innovation and New Technology*, 3, 183-199.
- Cohen, W. (1995). "Empirical Studies of Innovative Activity", in (eds) P. Stoneman, *Handbook of the Economics of Innovation and Technological Change*. Oxford, Oxford University Press.
- Crockett, G., P. Dawkins, P. Miller and C. Mulvey (1992). "The Impact of Unions on Workplace Productivity in Australia". *Australia Bulletin of Labour*, 18, 119-141.
- Deaton, A. (1997). *The Analysis of Household Surveys*. Washington, World Bank.
- Drago, R. and M. Wooden (1992). "The Australian Workplace Industrial Relations Survey and Workplace Performance". *Australian Bulletin of Labour*, 18, 142-167.
- DuMouchel, W. H. and G. J. Duncan (1983). "Using Sample Survey Weights in Multiple Regression Analyses of Stratified Samples". *Journal of the American Statistical Association*, 78, 535-543.
- Griliches, Z. (1986). "Productivity, R&D and Basic Research at the Firm Level". *American Economic Review*, 76(1), 141-167.

- Griliches, Z. (1995). "R&D and Productivity: Econometric Results and Measurement Issues", in (eds) P. Stoneman, *Handbook of the Economics of Innovation and Technological Change*. Oxford, Blackwells.
- Lichtenberg, F. R. (1993). "The Output Contributions of Computer Equipment and Personnel: A Firm-Level Analysis", NBER Working Paper 4540, National Bureau of Economic Research, Cambridge, MA.
- Mason, G. and D. Finegold (1997). "Productivity, Machinery and Skills in the United States and Western Europe". *National Institute of Economic Review*, 0(162), 85-98.
- Mohnen, P. (1992). *The Relationship between R&D and Productivity Growth in Canada and Other Major Industrialized Countries*, Ottawa: Canada Communication Group, Canada.
- Nadiri, M. (1993). "Innovations and Technological Spillovers", NBER Working Paper 4423, National Bureau of Economic Research, Cambridge, MA.
- Triplett, J. E. (1999). "The Solow Productivity Paradox: What Do Computers Do to Productivity?". *Canadian-Journal-of-Economics*, 32(2), 309-34.

Appendix: Survey data and regression analysis

Using regression techniques on survey data in economics is widespread. In many cases the survey data comes from a large scale survey, perhaps undertaken by a statistical service not directly linked to the researchers. These type of surveys often use stratified sampling (i.e. they break the population into separate ‘strata’ or groups and the sampling methodology varies across these groups). To use the example of the data used here, the Australian Bureau of Statistics have conducted the GAPS using strata based on firm size and industry. The reason for such methods is that stratified samples can produce more accurate estimates of population characteristics – such as mean profitability, or the proportion of firms innovating – for a given (overall) sample size (and implicitly cost of the survey). Within stratas the sampling methodology is often random, as is the case with the GAPS.

If the survey includes strata the data are normally provided with ‘weights’. To understand weights, let N_s be the number of observations in the population in the strata s , and n_s be the sample taken from strata s . Thus, the total sample size, n , is given by $\sum_s n_s = n$. The weight (w_i) is the ratio, N_s/n_s , and as such represents an ‘inflation factor’ or ‘scaling ratio’ for observations within strata. The sum of the weights equals the number of observations in the entire population (N) (i.e. $\sum_i w_i = N$, where i represents an observation in the sample).

Whether weights should be used in any analysis depends on the questions that a researcher aims to answer. To take the simplest example, assume that we are interested in the mean labour productivity across all firms in the GAPS data. Defining the population mean as \bar{p} , how can the data be used to produce an estimate of this? If the sample were random, not stratified, we would simply find the mean of the labour productivity in the sample. The strata, however, may create concern that the mean will vary across strata. In the case of labour productivity, this seems highly likely as small firms appear to have much lower labour productivity than large firms. Faced with this concern, we might calculate a separate mean for each stratum, \bar{p}^s . The population mean could then be calculated by

$$\sum_{s=1}^S \frac{N_s}{N} \bar{p}^s .$$

We can show that an equivalent method would have been to calculate a weighted mean using the weights provided in the data, i.e.

$$\bar{l}_w = \frac{\sum_{i=1}^n w_i l p_i}{N} = \frac{\sum_{s=1}^S \sum_{i=1}^{n_s} w_{si} l p_{si}}{\sum_{s=1}^S \sum_{i=1}^{n_s} w_{si}} = \frac{\sum_{s=1}^S \frac{N_s}{n_s} \sum_{i=1}^{n_s} l p_{si}}{\sum_{s=1}^S \sum_{i=1}^{n_s} w_{si}} = \frac{\sum_{s=1}^S N_s \bar{l}_s}{N} = \sum_{s=1}^S \frac{N_s}{N} \bar{l}_s .$$

This example shows that when we are concerned about heterogeneity in the population (i.e. the mean varying across strata) we can use a weighted mean to save the trouble of calculating means for all strata.

Consistency of OLS and WLS estimators under the assumption of Heterogeneity of coefficients

Consider now the issue of a multivariate regression model. Assuming that the vector of parameters differ across strata we can write:

$$y_s = X_s \beta_s + \varepsilon_s \quad s=1, 2, \dots, S.$$

where subscript s indicates that the vectors/matrices are stratum-specific. If the parameter of interest is the population-weighted β_p , following the concept of weighted means, β can be calculated by

$$\beta_p = \sum_{s=1}^S \frac{N_s}{N} \beta_s .$$

In practice, most researchers will estimate parameters by either OLS or WLS using the entire sample rather than estimate parameters by strata and calculate β_p . However, the following analysis shows that neither OLS nor WLS is a consistent estimator of β_p .

The OLS estimator for the entire sample can be written as

$$\beta_{OLS} = \left[\sum_{s=1}^S X_s' X_s \right]^{-1} \left[\sum_{s=1}^S X_s' y_s \right].$$

Again, assume the probability limit of the moment matrices are defined,

$$\text{plim}_{n_s \rightarrow \infty} \frac{1}{n_s} X_s' X_s = M_s \quad \text{and} \quad \text{plim}_{n_s \rightarrow \infty} \frac{1}{n_s} X_s' y_s = M_s \beta$$

The probability limit of β_{OLS} can be shown to be

$$\text{plim}_{n \rightarrow \infty} \beta_{OLS} = \left[\sum_{s=1}^S \frac{n_s}{n} M_s \right]^{-1} \left[\sum_{s=1}^S \frac{n_s}{n} M_s \beta_s \right]$$

The plim is therefore a weighted sum of the M_s and the $M_s \beta_s$. Assuming that as the sample size grows, the proportions in each stratum (n_s / n) are held fixed, β_{OLS} will be consistent for the common β only when the β_s is identical across strata.

A similar expression can be derived for the plim for the weighted estimator,

$$\text{plim}_{n \rightarrow \infty} \beta_{WLS} = \left[\sum_{s=1}^S \frac{N_s}{N} M_s \right]^{-1} \left[\sum_{s=1}^S \frac{N_s}{N} M_s \beta_s \right]$$

Again, the expression is a weighted sum of the M_s and the $M_s \beta_s$, only this time the weights are N_s/N , rather than n_s/n as in the OLS case. Whether the proportions of sample in each stratum are fixed or not is not an issue in terms the consistency since it is N_s/N , not (n_s/n) , in the expression. Again, this estimator is inconsistent for β_p if the β_s 's vary across strata; unless the M_s matrices are identical. It is very unlikely to have identical M_s in the real world. As Deaton (1997, p.70) stresses the problem of estimating the β is one of heterogeneity in the population, not the strata, so use of weights is of no use.



Minerva Access is the Institutional Repository of The University of Melbourne

Author/s:

ROGERS, MARK; TSENG, YI-PING

Title:

Analysing Firm-Level Labour Productivity Using Survey Data

Date:

2000-06

Citation:

Rogers, Mark and Tseng, Yi-Ping (2000) Analysing Firm-Level Labour Productivity Using Survey Data.

Persistent Link:

<http://hdl.handle.net/11343/33633>

File Description:

Analysing Firm-Level Labour Productivity Using Survey Data