Skills and productivity in the UK using matched establishment and worker data*

Jonathan Haskel

Queen Mary, University of London, Centre for Research into Business Activity (CeRiBA) and CEPR

Sonia Pereira

University College London and Centre for Research into Business Activity (CeRiBA)

17 March 2003

Abstract

The preliminary results in this paper show that high productivity is significantly associated with high levels of human capital measured in different ways. Results indicate that labour productivity in businesses in the top quintile of the productivity distribution have around 4.5 times higher productivity than businesses in the bottom quintile and around 90% percent higher measures of market value human capital. We also conclude that the experience part of human capital is less important in determining productivity than the education and other unobservable components of human capital.

---

* Contact: Jonathan Haskel, CeRiBA, Centre for Research Into Business Activity, Office for National Statistics, Zone D4-16, 1 Drummond Gate, London SW1V 2QQ, Email: jonathan.haskel@ons.gov.uk. This paper presents preliminary results, comments welcome. Work on this paper has been funded by a grant from the Evidence Based Policy Fund supported by HM Treasury, DTI and ONS, and carried out at CeRiBA at the Business Data Linking Branch at the ONS; we are grateful to all institutions concerned for their support. Results in this paper use the ONS ARD data set which is a research tool and may not entirely match official published figures.

CeRiBA Discussion Paper

Centre for Research into Business Activity –www.CeRiBA.org.uk

1 Introduction

This paper attempts to answer three questions:

a) do more productive establishments employ "better" workers;

b) if so, in what sense are they "better"; and

c) what fraction of variation in productivity is associated with variation in skills?

The main challenge in answering these questions is the lack of availability of establishment-level human capital and productivity measures. Typically, establishment level studies use rather crude measures of the establishments' workforce level of skills such as the ratio of production to non-production workers. Much work has used industry-level data often combining information on skills and productivity from various data sources. However, the skill variables usually used from household data – typically education and experience – capture only limited dimensions of skill. An 'ideal' data set would include worker level skills information on years and quality of education, advanced training, and personal skills matched with information on the employer's productivity.

In this paper we use matched employer-employee data using the New Earning Survey (NES) and the Annual Respondents Database (ARD)[1]. The ARD has establishment level information on outputs and inputs, from which we construct productivity (labour productivity and TFP). The NES does not have information on direct measures of the workers' education, qualifications or training. However, it does have information on occupation, age and wages, which in turn are correlated with workers' skills. Using the information available from the matched data, we can investigate if high productivity businesses differ from low productivity businesses in a number of skill dimensions. First, we can use occupational structure. Second, if wages reflect productivity we can see if high productivity businesses employ a mix of higher wage workers. Third, since we have a panel of employees matched with their employers, we derive human capital measures from panel regressions of wages on experience, person, establishment effects and other controls. Because we include establishment effects, we can use person effects and experience effects as a measure of worker human capital. Thus this measures worker skill, as

[1] The NES is a 1% sample of all workers with NI numbers. The ARD is a sample of businesses; in this paper we use the manufacturing data. The data are matched NES workers to an ARD establishment. Our NES data are available for 1994-6 and 1998-2000 (the establishment identifiers on the 1997 NES data are missing) and so our match is for these years. See Haskel and Pereira (2002) for details of the NES/ARD match and Barnes and Martin (2002) for more details of the ARD.

revealed in the market place, controlling for employer specific effects which can include a wide range of factors including employer specificwage policy.

2 Methods

2.1 Skill index from occupation code

We use Peter Elias (1995) skill classification which divides workers in four skills groups, based on the 2-digit occupation code, see Table 1.

Table 1

Skill Levels Based on the Standard Occupation Classification

Skill Level	Major Groups
Level 4	Managers and administrators (excluding office managers and managers/proprietors in agriculture and services) Professional occupations
Level 3	Office Managers and managers/proprietors in agriculture and services Associate professional and technical occupations Craft and relations occupations Buyers, brokers, sales reps
Level 2	Clerical, secretarial occupations Personal and protective service occupations Sales occupations (except buyers, brokers, sales reps) Plant and machine operatives Other occupations in agriculture, forestry, fishing
Level 1	Other elementary occupations

Source: Elias (1995)

Level 4 is the highest level of skill and includes managers, administrators and professional occupations (excludes office managers and managers/proprietors in agriculture and services). Level 3 includes office managers and managers/proprietors in agriculture and services, associate professional and technical occupations, craft and relations occupations, and buyers, brokers and sales representatives. Level 2 includes clerical and secretarial occupations, personal and protective service occupations, sales occupations (except buyers, brokers, sales representatives), plant and machine operatives, and other occupations in agriculture, forestry and fishing. Finally, level 1 is the lowest level of skill and includes other elementary occupations.

2.2 *Human capital measure from wage decomposition*

2.2.1 Wages and wage decompositions

In a competitive economy with perfect information, workers would be paid their marginal productivity at all times, and wages would therefore be a good measure of the market value of the workers' human capital. There are however various reasons why wages may differ from workers' marginal productivity. For example, in order to reduce labour turnover firms may pay lower wages in the beginning and higher wages latter in the job spell. Or, workers' bargaining power enables them to share some of the employer's rents through increased wages.

Wages may therefore depend on worker characteristics that affect their productivity as well as on firm characteristics, and in particular on the firm's wage policy, and the competitiveness of the market where the firm operates. Hildreth and Pudney (1998) classify workers as high and low-wage workers (and, by implication their employers as high- and low-wage firms) if their observed wage is above or below the expected wage predicted by their observable characteristics[2]. In other words, a worker is a low (high) wage worker if her observed wage is lower (higher) than the average wage of individuals with similar age, sex, hours of work, coverage by collective bargaining agreement, occupation and location. We build from their approach and estimate the expected wage by regressing (log) hourly real wages on the workers' characteristics as set out in (1):

$$\ln w_{it} = f_i + \alpha exper_{it} + \beta occ_{it} + \gamma X_{it} + Z_{J(i,t)} + \lambda_t + \varepsilon_{it} \qquad (1)$$

Where f_i is an individual fixed effect to capture individual unobserved time invariant differences in wages, $exper_{it}$ is a 4th degree age polynomial to allow for non-linear returns to labour market experience interacted with gender dummies to allow for different returns for males and females, occ_{it} is a vector with 2-digit occupation dummy variables, X_{it} is a vector including other variables with the workers' characteristics. It includes a dummy variable that is 1 for females and 0 for males and a dummy variable for full-time jobs. Z_{ji} is the vector of the variables of the establishment (industry, establishment size, region etc), and λ_t is a set of year

[2] They show how the residual wage (observed minus expected wage) correlates with performance indicators. Their results show that high-wage individuals tend to work for large employers with higher than average levels of profitability, investment, activity and sales. These individuals also have a lower probability than others of having entered their current job from the stock of registered unemployed.

dummies and ε a random error. From (1) we then construct various measures of human capital: the experience component, $\alpha exper_{it}$, the occupation component, βocc_{it}, the fixed unobservable labour quality f_i , and 'total' human capital h^1_{it} as defined in (2):

$$h_{it}^{1} = \alpha exper_{it} + \beta occ_{it} + f_{i} \tag{2}$$

With our matched data, however, we can go further than this. Because neither innate ability nor precise measures of education or further training are among the observed characteristics in our data, the residual wage component f_i includes all these[3]. However, since in (1) the characteristics of the employer are not accounted for, the residual wage component is also likely to include industry rents, employer specific wage policies, etc. To some extent this is controlled for in the X but we can exploit the fact we have matched data by using Abowd, Kramarz and Margolis et al (1999, 2001) decomposition of the (log) wages of individuals into a time-invariant individual affect, a time-invariant establishment effect and time varying observable individual characteristics. Formally, we regress the (log) hourly wage of individual *i* in business *J* at time *t* $(\ln w_{it})$ on a time invariant fixed effect f_i, a time invariant employer effect $\Psi_{J(i,t)}$, an experience polynomial interacted with gender dummy variables as before, $exp_{it}b$, a vector with 2-digit occupation dummy variables, occ_{it}, a vector X_{it} with two dummy variables, one for gender and one for full-time jobs:

$$\ln w_{it} = f_{i} + \alpha exper_{it} + \Psi_{J(i,t)} + \beta occ_{it} + \gamma X_{it} + \lambda_{t} + \varepsilon_{it} \tag{3}$$

where the function *J(i,t)* indicates the employer *J* of individual *i* at time *t*. The establishment effect is identified from workers' moves across establishments, and captures the component of the wage that is specific to the establishment. We therefore expect that the workers' individual effects will offer a much better approximation of the workers' human capital than measures derived from (1).

Identification of person and employer fixed effects requires that the employer has at least one worker with at least one job change; in a worker-business pair in which the worker has no other job spells observed in the data, one can not identify how much of the unobserved component in wages is due to the employer or to the worker. Furthermore, identification is problematic across

[3] Note that f_i is by definition fixed over time.

different groups[4] of connected individuals and employers (Abowd, Creecy and Kramarz, 2002). In fact, because individual fixed effects are estimated by including a set of person dummies in (2) and establishment effects by including a set of business dummies, the two full sets of dummies and the constant term are not uniquely identified. One possible way to proceed is to set both the mean on the individual effects and the mean of establishment effects equal to zero. However, because the data set includes various separate groups of connected individuals and employers, this identifying restriction would have to be applied to each group of connected individuals and employers. Given that in our data, there are many small groups, setting the means on the individual effects and the of establishment effects equal to zero for each of these small groups can be problematic, since some groups can comprise workers with vary different average skills and/or very different average establishment characteristics. In fact, if this is the case individual and establishment effects are not comparable across groups of connected individuals and employers. Therefore, in this paper, because of the nature of our data we estimate model (3) on the largest group of connected individuals and employers (see appendix for details) rather than for all possible groups as Abowd et al do. We then use the (log) wage decomposition in (3) as per (2) i.e. construct a measure of human capital similar to the one used in Abowd et al (2002), which is the worker fixed effect plus the contribution of time varying observable individual characteristics (experience) to the worker's wage. Note that one limitation of this technique is that the estimation of fixed effects by definition only captures the workers' wage component that is fixed over time. This will of course not capture the heterogeneous accumulation of human capital over time (human capital is allowed to increase overtime with experience, but equation (3) assumes that the impact of years of experience – proxied by age – is similar for all individuals with the same gender).

2.2.2 Productivity

For manufacturing labour productivity is constructed at the business unit level and is measured as real gross output per worker. We also have measures of real capital per worker and real material use per worker. These are fairly standard measures from the ARD.

[4] A group of connected persons and employers contains all the workers who are observed to have ever worked for any of the employers in the group and all the employers at which any of the workers were ever employed. The groups of connected individuals and establishments can be determined by applying methods from the graph theory (Abowd, Creecy and Kramarz, 2002).

3 More productive establishments and "better" workers?

Table 2 shows mean and median wages and productivity by productivity quintile. The productivity distribution for each year is divided in five quintiles, and matched businesses are grouped into their relevant quintiles[5]. Quintiles (column 1) are presented by increasing order in productivity. Column 2 shows that there are around 4,320 businesses in each quintile, and column 3 shows that the number of matched workers varies from 9,854 in the lowest quintile to 23,457 in the top quintile, reflecting the relationship between size and productivity. Column 4 confirms that the total number of workers employed in the businesses matched in the top quintile (quintile 5) is over twice as high than the total employment in the bottom quintile (quintile 1).

Table 2

Mean and median wages and productivity by productivity quintile

Productivity quintile	Nbr. Businesses	Nbr. Matched workers	Total employment in the businesses matched	Percent difference of quintile *mean labour productivity* from first quintile	Percent difference of quintile *median labour productivity* from first quintile	Percent difference of quintile *mean wages* from first quintile	Percent difference of quintile *median wages* from first quintile
1	2	3	4	5	6	7	8
1	4,317	9,854	949,899	0.000	0.000	0.000	0.000
2	4,321	12,238	1,205,644	0.540	0.463	0.152	0.183
3	4,320	14,228	1,489,834	1.052	0.941	0.250	0.292
4	4,321	14,906	1,573,584	1.856	1.683	0.335	0.393
5	4,323	23,457	2,419,250	4.374	3.742	0.496	0.585

Source: Authors' calculations from full matched employer-employee data set.

Columns 5 and 6 show the mean and median productivity relative to the bottom quintile. These are the difference between each quintile mean (median) labour productivity and the bottom quintile mean (median) productivity divided by the latter. This was done for each year and averaged across all years. While the mean labour productivity (column 5) in the 4th quintile is nearly twice as high as labour productivity in the bottom quintile, labour productivity in the 5th

[5] Because our worker data covers 1 percent of the workers, in most cases we only have information for a small sample of workers in each business. In fact, in many business we just have one worker matched. Given that for most businesses the number of workers matched is very small, we cannot undertake a meaningful business level analysis. By partitioning the data into 5 productivity quintiles, we have however plenty of matched workers to investigate whether there is a correlation between productivity and wages.

(top) quintile is over 4 times higher than labour productivity in the bottom quintile[6]. This suggests that the labour productivity distribution has a long upper tail in relation to its lower tail (right skewed distribution), which in turn suggests a higher heterogeneity in terms of labour productivity among the most productive establishments. In fact, though relative median productivity (column 6) is lower than relative mean productivity (column 5) for all quintiles and this difference is markedly larger for the 5th quintile. Columns 7 and 8 show the mean and median hourly wages relative to the bottom quintile. Mean and median wages increase with the productivity quintile, suggesting that on average wages are higher in more productive businesses. However, the wage gap between the top and bottom productivity quintiles is remarkably lower than the corresponding productivity gap. For example, mean wages in the 5th quintile (column 7) are roughly 50 percent higher than mean wages in the bottom quintile. Relative median wages (column 8) are slightly higher than relative mean wages (column 7), suggesting that the wage distribution is slightly skewed to the left.

Table 3 shows the shares of the skill levels across quintiles, with skill measured by the occupational mix as set out in Table 1. The table shows that lower productivity quintiles employ higher shares of low skilled workers (column 2) and lower shares of high skilled workers (column 5). In addition, column 3 shows that most productive businesses employ more of skill level 4 workers (column 5) and less of skill level 3 workers (column 4) than less productive businesses.

Table 3

Share of skill levels of productivity quintiles (skills measured by occupation)

Prod'y quintile 1	Share of skill level 1 2	Share of skill level 2 3	Share of skill level 3 4	Share of skill level 4 5
1	0.046	0.468	0.398	0.088
2	0.050	0.508	0.344	0.099
3	0.048	0.499	0.327	0.126
4	0.039	0.508	0.310	0.143
5	0.038	0.492	0.291	0.179

Source: authors' calculations from full matched employer-employee data set. Number of workers and businesses in table 2.

Table 4 presents a four measures of relative human capital computed from equation (1) estimates as follows:

[6] The data was carefully screened for outliers, and the top and bottom percentiles of the labour productivity distribution were dropped to reduce the possibility of remaining outliers. The results are

$$\text{person fixed effects} = \frac{\sum_{i=1}^{n_q} f_i^{Q=q}}{n_q} - \frac{\sum_{i=1}^{n_1} f_i^{Q=1}}{n_1}$$

$$\text{experience component} = \frac{\sum_{i=1}^{n_q} exper_{it}^{Q=q}\alpha}{n_q} - \frac{\sum_{i=1}^{n_1} exper_{it}^{Q=1}\alpha}{n_1}$$

$$\text{occupation component} = \frac{\sum_{i=1}^{n_q} occ_{it}^{Q=q}\beta}{n_q} - \frac{\sum_{i=1}^{n_1} occ_{it}^{Q=1}\beta}{n_1}$$

where Q refers to the productivity quintile and n_q is the number of businesses in quintile q and where human capital = person fixed effects+ experience component +occupation component.

Each of the above these measures are computed for each year and averaged across all years. We shall express percentage differences in these measures relative to the lowest quintile. Note that because these measures are obtained from the log-wage equation in (1), they are approximations of percent differences of the various wage components between each quintile and the bottom quintile. Table 4 shows the results for the exact percent difference which are obtained from applying the exponential transformation to the four above measures and subtracting 1.

Results in columns 2 to 5 are on the basis of equation (1) but with no establishment regressors. They suggest that high productivity is associated with higher human capital for all measures of human capital: unobserved individual fixed effect (portable skills not measured by the other variables), experience component of wages, occupation component and the sum of the three. Columns 6 to 9, which include establishment regressors in (1), show that this result holds after including employer characteristics (2-digit industry dummies, 2-digit industry dummies interacted with time dummies, size of the employer, and region).

Table 4

Skill measures derived from equation (1) across productivity quintiles: for whole sample

	No establishment controls in (1)	Establishment controls in (1)
Productivity quintile	Person Fixed Effequh7.8696 T 0 048 5.762 Tc(E)386.7(u)2Exp2(ti-13(3s)11.6(t-19.112619 4(t)-ce5 -1.2382143-0.0473 Tc(E)-5.7(u)2Co Fixed Effequh7.8696 T 0473 5.762 Tc(E)386.7(u)2Exp2(ti-13(3s)11.6(t-19.112619 4(t)-ce5 -1.2382143-0.0473 Tc(E)-5.7(u)2Co.	

1	2	3	4	5	6	7	8	9
1	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
2	0.101	0.166	0.010	0.296	0.092	0.166	0.008	0.283
3	0.162	0.252	0.020	0.484	0.146	0.252	0.017	0.460
4	0.234	0.241	0.024	0.569	0.212	0.242	0.021	0.538
5	0.374	0.241	0.024	0.747	0.336	0.241	0.022	0.696

Source: authors' calculations from full matched employer-employee data set. Number of workers and businesses in table 2.

Table 5 repeats the regression (1) analysis but for the smaller sample of establishments on which person and establishment fixed effects can be identified. We do this to ensure comparability with estimates of equation (3) which are shown in table 6. In spite of the considerable difference in sample size, table 5 shows a very similar pattern to table 4. We however note that after including establishment controls (columns 8 to 11) the relative difference in the person fixed effects between the top and bottom quintiles is 41.2 percent in the smaller sample and 33.6 percent in the full matched sample. In addition, the relative difference across quintiles in the experience and occupation components in the full matched is nearly double the one in the smaller sample.

Table 5

Skill measures derived from equation (1) across productivity quintiles: for largest group of connected establishments for whom person and establishment fixed effects are identified.

			No estab. Controls in (1)				Establishment controls in (1)			
Productivity quintile	Nbr. Businesses	Nbr. Matched workers	Person fixed effects	**Exper. compo-nent**	**Occup. Compo-nent**	Human capital	Person fixed effects	**Exper. compo-nent**	**Occup. Compo-nent**	Human capital
1	2	3	4	5	6	7	8	9	10	11
1	838	4,435	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
2	841	4,668	0.092	0.312	0.007	0.443	0.095	0.308	0.003	0.440
3	840	4,891	0.151	0.320	0.013	0.543	0.149	0.320	0.006	0.537
4	841	5,777	0.316	0.363	0.024	0.834	0.302	0.361	0.010	0.806
5	842	8,506	0.427	0.348	0.026	0.975	0.400	0.346	0.012	0.926

Source: authors' calculations from employer-employee data set for largest group of connected establishments for whom person and establishment fixed effects are identified. Number of workers and businesses in table 2.

Finally, table 6 shows the results of using skill measures obtained from (3) i.e. the equation with both establishment and person fixed effects on the sample for which this is feasible. Again, high productivity is associated with high human capital levels. According to the measure of human capital computed in (2), businesses in the top quintile have on average 87.5 percent higher human capital than businesses in the bottom quintile. This number is much smaller than the corresponding numbers obtained with regressing equation (1) (62.1 percent in table 5, column11, and 49.7 percent in table 4, column (9). This is the result we expect since the establishment fixed effects included in equation (3) are likely to capture a much greater component of the establishment specific wage policy/ wage rents, etc. that the establishment controls in equation (1).

Table 6

Skill measures derived from equation (3) across productivity quintiles: for largest group of connected establishments for whom person and establishment fixed effects are identified.

Prod. Quintile	Nbr. Businesses	Nbr. Matched workers	Total employment in the businesses matched	Percent difference of quintile *mean labour productivity* from first quintile	Percent difference of quintile *mean wages* From first quintile	Establish ment fixed effects	**Person Fixed Effects**	**Exper. compo-nent**	**Occup. Compo-nent**	**Human capital *H*[2]**
1	2	3	4	5	6	7	8	9	10	11
1	838	4,435	494,618	0.000	0.000	0.000	0.000	0.000	0.000	0.000
2	841	4,668	561,596	0.576	0.146	0.117	0.074	0.367	-0.001	0.473
3	840	4,891	558,333	1.138	0.225	0.136	0.122	0.380	0.007	0.566
4	841	5,777	688,346	2.019	0.421	0.230	0.192	0.434	0.011	0.723
5	842	8,506	1,045,338	4.590	0.543	0.198	0.315	0.425	0.012	0.897

Source: authors' calculations from employer-employee data set for largest group of connected establishments for whom person and establishment fixed effects are identified. Number of workers and businesses in table 2.

These results suggest two conclusions to the questions set out in the introduction. First, more productive establishments hire "better" workers. Second, the dimensions of "better" are from an occupationally higher skill mix, better paid and having better personal characteristics, controlling for age and employer, as rewarded in the market-place. More productive establishments do not employ workers whose experience is rewarded, controlling for other things, but they do employ better workers in the sense that they attract higher waged workers at that establishment, controlling for their other factors.

4 How much more productive are establishments with "better" workers?

To explore this we run the regression

$$\ln(Y/L)_{Jt} = \alpha_1 \ln(K/L)_{Jt} + \alpha_2 \ln(M/L)_{Jt} + \alpha_3 SKILL_{Jt} + \lambda_t + \lambda_I + \varepsilon_{Jt} \qquad (4)$$

where ln(Y/L) is log gross real output per worker in establishment J, ln(K/L) is log capital per worker in J, log(M/L) is log real materials use per worker in J, SKILL is the establishment-specific skill measure (to be detailed below), λ_I and λ_t are 5 digit industry and time dummies.

To implement this regression we need to calculate skill for each establishment. We did this by taking all establishments on the matched employer-employee sample for whom we could run the regression (3). This gives a set of person-specific skill measures, namely the person effects, f_I, the experience component and the human capital effect. We then calculated the mean of these effects for each establishment. Of course, for some establishments we only have one matched worker, and so we included only establishments for whom we had at least 10 matched workers (the median number of matched workers per establishment was 13).

Following Abowd et al (2002), we calculated the following skill indices. For each establishment, we calculated the share of workers whose human capital, h as in (2), is greater than overall median h. We then did the same for the person-effect and experience component. The results of this exercise are set out in Table 7.

Table 7

Productivity and skills: estimates of (4) (dependent variable: $\log(y/l)_J$)

	(1)	(2)	(3)	(4)
Average Hcap	0.0763			
	(0.0207)**			
Average Person f_i		0.1002		
		(0.0252)**		
Average Exper		-0.0017		
		(0.0508)		
Fraction of workers with Hcap above median			0.0665	
			(0.0496)	
Fraction of workers with Person f_i above med.				0.1227
				(0.0386)**
Fraction of workers with Exper above median				-0.0925
				(0.0768)
LnK/L	0.0311	0.0304	0.0367	0.0337
	(0.0126)*	(0.0126)*	(0.0129)**	(0.0127)**
lnM/L	0.6919	0.6935	0.6981	0.6917
	(0.0175)**	(0.0174)**	(0.0176)**	(0.0177)**
Observations	685	685	685	685
R-squared	0.9237	0.9240	0.9221	0.9067

Note: Standard errors in parentheses, based on robust standard errors, * significant at 5%; ** significant at 1%. Other controls are 5-digit industry dummies, and year dummies.

We have 685 establishments in our sample. Column 1 shows that establishments employing workers with a higher average human capital have higher productivity. Column 2 splits the human capital effect into the person and experience components and shows that the positive human capital effect arises mostly from the person effect. Columns 3 repeats the analysis using as measure of human capital the share of workers with human capital above that year's median. The coefficient is not very different from the result with mean human capital (column 1), though it is now insignificantly different from zero. Column 4 shows that it is the share of workers with person effect above that year's median that matters, and not the experience component.

Finally to get some idea of how much changes in skills change productivity, column 1 of Table 6 suggest that a one standard deviation change in human capital measure (sd=0.66) gives a 0.05 (=0.66*0.0763) change in log labour productivity. Given that the standard deviation of log labour productivity is 0.77 this is about 6% of the variation (note however that a one standard deviation change in log M/L gives a 0.61 change in log labour productivity). For comparison, Abowd et al (2002) find the contribution of a one standard deviation change in h of 0.19 to log labour productivity, with log labour productivity having a standard deviation of 0.84 (although their gross output regressions do not include a materials term). However, their regression has the dependent variable of gross output, but does not include materials, only capital.

5 Conclusions

The preliminary results in this paper show that high productivity is significantly associated with high levels of human capital measured in many different ways. Results indicate that labour productivity in businesses in the top quintile of the productivity distribution have on average 4.374 times higher productivity than businesses in the bottom quintile and between 30 and 70 percent higher measures of market value human capital. We also conclude that the experience part of human capital is less important in determining productivity than the education and other unobservable parts of human capital.

6 References

Abowd, J., Creecy, R. (2002), 'Computing Person and Firm Effects Using Linked Longitudinal Employer-Employee Data', mimeo.

Abowd, J., Haltiwanger, J., Jarmin, R., Lane, J., Lengermann, P., McCue, K., McKinney, K., Sandusky, K. (2002), 'The Relationship Between Human Capital, Productivity and Market Value: Building Up from Micro Evidence', mimeo.

Abowd, J., and Kramarz, F., (1999), 'The Analysis of Labour Markets Using Matched Employer-Employee Data', in O. Ashenfelter and D. Card, (eds), Handbook of Labour Economics, Norht Holland, pp. 2629-2710.

Barnes, M. and Martin, R., (2002), 'Business Data Linking: An introduction', *Economic Trends*, 581, April, pp. 34-41.

Elias P. (1995), 'Social class and the Standard Occupational Classifications', in D.Rose (ed.), A report on phase 1 of the ESRC Review of the OPCS Social Classification, Economic and Social Research Council: Swindon.

Haskel, J., and Pereira, S., (2002), "Skills and productivity: first results from the matched NES/ARD data set", draft note.

Hildreth, A. and Pudney, S. (1998a), 'Understanding the Labour Market: Matching Workers and Employers using Respondent-Level Data', Labour Market Trends, pp 643-656.

Hildreth, A. and Pudney, S., (1998b), 'Econometric Issues in the Analysis of linked Cross-section Employer-worker Surveys', mimeo,
available at http://www.le.ac.uk/economics/research/RePEc/lec/lpserc/pserc98-3.pdf.

Hildreth, A. and Pudney, S., (2000), 'Econometric Issues in the Analysis of linked Longitudinal Employer-Worker Data', mimeo.

7 Appendix

Table A1 shows the correlation coefficient for the components of log real hourly wages for different estimated equations. There is a strong positive correlation between the log wage and the person effect with less strong between the log wage and the establishment effect. Like in other work, there is a negative correlation between the person and the establishment effect.

Table A1

Correlation coefficients for components of log real hourly wage

				Largest group no establishment fixed effects			Largest group establishment fixed effects			
			(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
			log real wage	Indiv. Effect	exper. comp.	human capital	indiv. effect	exper. comp.	human capital	establish ment effect
	(1)	log real wage	1.000	0.738	0.357	0.948	0.525	0.149	0.592	0.154
Largest group no establishment fixed effects	(2)	individual effect	0.738	1.000	-0.288	0.779	0.591	-0.322	0.490	0.105
	(3)	experience component	0.357	-0.288	1.000	0.377	-0.027	0.716	0.229	0.061
	(4)	human capital	0.948	0.779	0.377	1.000	0.554	0.157	0.624	0.142
Largest group establishment fixed effects	(5)	individual effect	0.525	0.591	-0.027	0.554	1.000	-0.242	0.938	-0.664
	(6)	experience component	0.149	-0.322	0.716	0.157	-0.242	1.000	0.110	0.008
	(7)	human capital	0.592	0.490	0.229	0.624	0.938	0.110	1.000	-0.677
	(8)	establishment effect	0.154	0.105	0.061	0.142	-0.664	0.008	-0.677	1.000

Note: Human capital (rows and columns 4 and 7) is defined as the sum of the experience component (rows and columns 3 and 6) and the estimate of the unobserved individual fixed effect (rows and columns 2 and 5).