Economic Issues, Vol. 7, Part 2, 2002

Assessing the Determinants of Male Earnings Dispersion

Karl Taylor[1]

Abstract

This paper considers male earnings dispersion in the United Kingdom in four industries from 1973 to 1995. The analysis takes place in two stages. Firstly, earnings dispersion over time is split into two components: between-group earnings dispersion due to differing worker characteristics across the population; and within-group earnings dispersion, that is any remaining earnings dispersion after controlling for measurable worker characteristics. Secondly, that part of earnings dispersion which cannot be explained by observable worker characteristics is examined by industry using time series techniques to assess the impact of technological change; globalisation; female participation; immigration; and institutional changes upon remaining dispersion.

1. Introduction

The primary objective of this paper is to examine developments in the British wage structure over the period 1973 to 1995. Evidence to date on earnings dispersion has either been based on the economy as a whole (Schmitt, 1995) or, for disaggregated manufacturing industries (Machin, 1996a,b). This study focuses upon four industries for reasons of consistency over time (see section 4). These are Manufacturing, Other Manufacturing, Construction, and Transport and Communication. Of particular interest is to assess whether the four industries experienced the same trends in earnings dispersion. By considering specific industries outside of the manufacturing sector, it is possible that different factors have played a significant role in each industry, once the effects of worker characteristics have been controlled for. For instance, the role of female participation and immigration - supply side pressures, may be of greater importance than demand influences (technological change, and globalisation) in certain industries. Also factors contrary to the market mechanism may be at work, specifically declining collective bargaining as unionisation falls.

2. Background to earnings dispersion

Over the past two decades the gap between the richest and poorest members of society has widened (Goodman *et al.*, 1997). Earnings are an important part of overall income, and the trend in the dispersion of earnings closely follows the trend in the dispersion of overall income (Gosling *et al.*, 1996). Previous research has shown that earnings dispersion fell during the

1970s, only to increase rapidly during the 1980s (Schmitt, 1995). Part of the change in earnings dispersion can be related to changing returns to labour market skills such as education and experience. It is possible to disaggregate earnings dispersion into between-group and within-group components. Between-group dispersion accounts for earnings dispersion arising due to different levels of characteristics amongst individuals. For example, between the young and old, the highly educated and the school leaver with minimal qualifications, low experienced and high experienced individuals, whites and non-whites, regional pay variations and wage differentials between industries. The sharp rise in earnings dispersion between education and experience groups, manifests itself in the substantial growth in the financial returns to education and experience that took place in Britain during the 1980s (Schmitt, 1995). However, earnings dispersion has also increased within specific groups defined by characteristics such as age, education, experience, colour, region and industry. Over the past two decades between-group effects have explained only a portion of the overall rise in earnings inequality (Schmitt, 1995 and Machin,1996a).

The majority of the trend in increasing earnings dispersion, can be explained by a widening distribution of earnings occurring within-groups of workers, possessing similar experience and educational characteristics (Schmitt, 1995; Machin, 1996a). Any explanation of rising earnings dispersion must be capable of accounting for these within-group changes. A shift in relative labour demand in favour of workers with high levels of skills appears to be the most likely explanation (Levy and Murnane, 1992; Gosling *et al.*, 1996). This increase in earnings dispersion occurring within narrowly defined groups has been related to several factors, including changes in demand and supply patterns for labour, and the influence of pay setting institutions. To the extent that unions have maintained reasonable levels of pay by creating wage floors above the market clearing level, a marked decline in collective bargaining can be expected to have influenced earnings dispersion. The key themes apparent in the literature which explain the collapse in the demand for lower skilled male workers are now briefly introduced, along with the role of institutional changes.

A number of explanations exist to explain this relative demand change, the two most common being skill-biased technological change (Krueger, 1993; Haskel, 1999; Autor, Katz and Krueger, 1998; Machin and Van Reenen, 1998) and the growth in international trade (Wood, 1994, 1998; Anderton and Brenton, 1999). Both the expansion of technology and growth in international trade can be considered as crucial elements of recent globalisation.

The evidence that skill-biased technological change has increased demand in favour of skilled labour is twofold, both indirect and direct. Indirect evidence has relied upon residual wage inequality from standard earnings functions, whilst direct evidence is based upon correlations between wage differentials and indicators of technological progress. The indirect evidence is open to criticism since empirically technical change is typically defined to be the amount of change in relative wages that cannot be explained by observable characteristics, that is the residual, which arguably could also be explained by globalisation, supply side factors or institutional changes. More recently, direct evidence has provided a link between indicators of technology, such as computer use or research and development intensity, and wage inequality (Krueger, 1993; Autor *et al.*, 1998; Machin and Van Reenen, 1998).

Globalisation arguments suggest that developed countries have become increasingly open to competition from lower wage developing economies. Consequently, firms have taken the opportunity to gain from these lower costs by substituting unskilled intensive production abroad. A number of authors have recently argued that such outsourcing is important in explaining wage inequality (Wood, 1994, 1998; Anderton and Brenton, 1999; Feenstra and Hanson, 1999).

Less common explanations apparent in the literature, which focus on market forces, are the role of female participation and immigration. Both of these factors may increase the supply of relatively low skilled labour, and thus drive down the wages of low skilled workers. Alternatively, the impact of both changing female participation rates, and immigration is largely dependent upon the degree of substitutability for low skilled males. For example, if females or immigrants are substitutes for low skilled workers, then a rise in the supply of either leads to a fall in the demand for the lower skilled (Topel, 1997). The substitution argument is more appealing due to the well-documented demand shift in favour of skilled labour and the fact that the supply of skilled labour has increased (Schmitt, 1995), aided by female participation (Harkness, 1996) and immigration (Bell, 1997).

Aside from market force explanations, other authors have stressed the importance of labour market institutions, in particular trade unions, in shaping the way labour markets have responded to these changes in demand and supply (Freeman, 1993; Gregg and Machin, 1994; and Machin, 1997). Market force explanations can explain many of the similarities in the development of the wage structure, but are less illuminating when attempting to explain differences (Gottschalk and Smeeding, 1997). Most economies have been subjected to increased technological change and globalisation, yet only the United Kingdom and United States experienced substantial increases in earnings dispersion (Katz, Loveman and Blanchflower, 1995). One could argue that it is such labour market institutions that explain differing wage inequalities across countries. Following the same logic different institutional changes across industries over time may account for some of the trend in earnings dispersion.

An innovative approach used in the following analysis is the two-stage empirical methodology adopted to analyse earnings dispersion. Initially, repeated cross sections of the annual *General Household Survey* (*GHS*) over a period of 23 years are used to control for differences in earnings, which may arise between individuals stemming from differences in experience, education, colour and region. This enables earnings dispersion to be split into between-group and within-group components, following Schmitt (1995) and Machin (1996a). The between-group component is explained by data available from the *GHS* based upon individuals in the population, and arises due to changing returns to individual characteristics. Of potentially greater importance is the trend in within-group earnings dispersion over the 23 years, that is what cannot be explained by the micro data. In the past any remaining dispersion after controlling for individuals characteristics has usually been attributed to technological change (Gottschalk and Smeeding, 1997). However, in this approach the second stage employs time series analysis to consider the role not only of technology, but other possible influences upon unobserved dispersion namely globalisation, female participation, immigration and labour market institutions. Of particular interest is how each may have influenced the trend in within-group earnings dispersion over time in each of the four industries considered.

In the absence of panel data, a two-stage empirical methodology is deemed preferable to alternatives for a number of reasons. First, there would be problems of pooling the data over time, because data upon individuals is used along with more aggregate industry level data.

Consequently, pooling could result in aggregation bias where estimates are downwardly biased (Moulton, 1986). Even if this problem is overcome by the use of cell-means or corrected standard errors there is a possibility that the industry level data may be non-stationary over time. This presents a major problem, in that data-pooling without considering the stationarity of the variables, can result in a spurious correlation. The two-stage approach draws together different strands in the literature. Previously earnings dispersion has been decomposed into between-group and within-group dispersion using individual data (Juhn et al., 1993; Schmitt, 1995; and Machin, 1996a). Time series methods have been used to study the potential causes of earnings dispersion over time (Borjas and Ramey, 1994; Buckberg and Thomas, 1996; and Leslie and Pu, 1995, 1996), but only at the economy wide level, and not for a measure of earnings dispersion purged from differing returns to worker characteristics. The empirical approach used in this study combines these two approaches. First, earnings dispersion is purged from measurable differences in the distribution of education, experience and personal characteristics across the population, all of which may affect the trend in earnings dispersion. Second, a time series approach is taken to avoid the problems of aggregation bias and non-stationarity, and examine the major contributor to within-group earnings dispersion for each industry.

Research to date has only offered snapshots for particular years,[2] rather than forming a consistent time series of within-group earnings dispersion. Whilst snapshots allow the analysis of earnings dispersion between two static periods of time, they are less informative about the trend of earnings dispersion over time. In the absence of large scale, long term panel data sets, repeated cross sections offer the best insight available, into the structural changes that have occurred in the British labour market over the last two decades.

Section 3 introduces the empirical methodology used to decompose earnings dispersion into between-group and within-group components, and the time series methods used to assess the importance of the dominant themes in the literature upon within-group earnings dispersion. Section 4 considers the data required to undertake the analysis, followed by a presentation of the results in Section 5.

3. Empirical methodology

The empirical framework takes place in two steps. Firstly micro data based upon the individual are used to control for differences across the population in experience, education, personal characteristics and regional location - all of which may influence earnings. This enables earnings dispersion to be split into within-group and between-group components, following Juhn et al. (1993) and Machin (1996a). In the second step time series techniques are adopted, based upon aggregate industry data attempting to proxy market forces and institution change, in an attempt to explain the trend in within-group earnings dispersion over time, based upon the methodology of Borjas and Ramey (1994). The following introduces the empirical approach used to evaluate the possible determinants of within-group earnings dispersion and test the following hypotheses:

H_0: After controlling for measurable worker characteristics, no factor analysed in the industry level data can explain the trend in within-group earnings dispersion.

H_1: There has been a decline in the relative demand for lower skill endowed workers, that is demand factors in particular skill biased technological change and globalisation are responsible for within-group earnings dispersion.

H_2: Alternatively supply side influences may be the cause of within-group earnings dispersion, where changes in labour supply are alleged to have exacerbated dispersion particularly the changing patterns of immigration and female participation rates. The impact of such supply side factors upon within-group earnings dispersion within particular industries depends upon substitution possibilities, for instance whether women and/or immigrants who have entered the labour market are good substitutes for low skilled males. Alternatively, either females or immigrants may just increase the supply of the unskilled, hence driving down the price of lower skilled workers.

H_3: The decline in unionisation and changes to the structure of collective bargaining is the prevailing factor in determining within-group earnings dispersion.

Decomposing earnings dispersion

One problem with existing studies is that the measure of dispersion used is typically a ratio of one relatively skilled group to a less skilled group, this raises the issue that any inference about determinants of dispersion assumes an equal distribution of human capital characteristics amongst groups of individuals. This is unlikely, and the approach taken here compensates for this by deriving a measure of within-group earnings dispersion by industry free from the influence of measurable worker characteristics. A regression framework is used to control for specific individual characteristics (given as the vector **X** in equation 1, below) such as experience, employment status, colour, marital status, education and regional location. Under such a scenario within-group earnings dispersion can be seen as the dispersion of the residual from the regression (Juhn *et al*., 1993), where a wider dispersion of the residuals shows greater earnings dispersion occurring within-groups. Such dispersion is important to understand, as the majority of earnings dispersion occurred within narrowly defined groups in the UK - Schmitt (1995) and Machin (1996a). The regressions are estimated cross sectionally by industry over the period gaining a measure of within-group earnings dispersion over two decades.

$$Log(Wages)_i = \alpha_i + \mu_1 Exp_i + \mu_2 Exp_i^2 + \lambda Colour_i + \kappa Marital\ status_i + \psi Job\ status_i$$
$$+ \phi \boldsymbol{Education}_i + \gamma \boldsymbol{Region}_i + \varepsilon_i \equiv \boldsymbol{X}_i \delta + \varepsilon_i \qquad \forall_{j,t}$$
$$\varepsilon_i \sim IID(0, \sigma^2) \tag{1}$$

The interpretation given to the residual ε_i from an equation based upon the above, is that after controlling for personal characteristics, human capital endowments and regional location, its standard deviation represents within-group earnings dispersion. Thus, within-group inequality in the j^{th} industry at time t is given by the standard deviation of the residual and is that part of wages which can not be explained by worker characteristics:

$$v(\hat{\varepsilon}) = \sqrt{\sum_{i=1}^{n} \left[\left(\hat{\varepsilon}_i - \hat{\bar{\varepsilon}} \right)^2 / (n-k) \right]} \qquad \forall_{j,t} \tag{2}$$

where a 'hat' indicates the coefficient is an estimate. Thus we now have a scalar measure of within-group earnings dispersion for each industry and time period. In the past residuals from earnings functions such as equation 1 have widely been attributed to skill biased technological change. Note that traditionally the residual from a standard wage equation was seen as indirect evidence of skill-biased technological change (Katz and Autor, 1999). However, there is no reason why the other factors mentioned in section 2 could not also explain part of the remaining residual. This paper moves beyond the assumption that the residual represents technological change by assessing other potential influences.

Explaining trends in within-group earnings dispersion

The second stage of the analysis uses time series techniques to consider the impact of the market forces and institutional changes on within-group earnings dispersion for each industry. Data on demand changes, supply changes and institutional change are required, and initially these data and the measure of within-group earnings dispersion derived from step one are checked for stationarity. Stationarity is an important concept because the standard regression model makes assumptions about the stationarity of the disturbance term as well as the stationarity of the variables in the regression. In particular the standard regression model assumes that the errors are drawn independently from a white noise process and that the independent variables are random stationary processes independent of the residual. A variable with a stochastic trend is a typical case of a non-stationary process, where regressing a mixture of trended and non trended variables against one another is likely to result in a spurious regression. Thus initially the market force indicators, institutional change and within-group earnings dispersion are tested for stationarity, using two alternative techniques.

The most common test used for checking the stationarity of data is the Augmented Dickey Fuller (ADF) test (Dickey and Fuller (1979)) which allows for deterministic components in the form of a constant and trend. The null hypothesis is that the variable is non-stationary. A particular problem with the ADF test is that the results may be distorted because of the small sample size i.e. 1973 to 1995 gives 23 observations. Kwiatkowski *et al.* (1992 - hereafter KPSS) propose an alternative test for unit roots, where the asymptotic validity of the test holds for fairly small sample sizes and if the lag length is set to zero tests will not be subject to size distortion. Furthermore, instead of the null hypothesis of non-stationary data, the null hypothesis is reversed to the case of stationary data, and so will not be rejected unless there is strong evidence in contrast to the usual case of non-stationary data. Both the ADF and KPSS approaches are used for testing the stationarity of the data.

Much of the literature has sought to explain fluctuations in wage relativities by analysing data that has been first differenced or detrended. However this type of analysis removes the trend component, where clearly the long term persistent movements of the trend in relative wages is of importance. By first differencing data researchers are only analysing year-to-year growth rates. The argument made here is that the best way to actually proceed is to analyse the levels of the relevant variables, rather than their differences. A possible problem that may be envisaged of the two step approach is that for the second step we only have twenty-three observations, yet time series models typically use hundreds of observations and Reimers (1992) suggests that by using small sample sizes the null hypothesis of no cointegration is over rejected. Parker (2000) has also raised doubts about the power of cointegration tests when sample sizes are small, particularly in the context of inequality. However this can be compensated by taking into account the number of parameters to be estimated and making an adjustment for the degrees of freedom when computing test statistics. Also despite the small number of observations it has been argued that the actual span of the data is more crucial in determining a tests power rather than the specific number of intervals or data frequency, Maddala and Kim (1998) and Otero and Smith (2000).

The cointegration approach considers whether wage dispersion follows the same trend as its potential causes. A particular advantage of cointegration analysis is that it can be used to assess what has influenced dispersion over time, by assessing the magnitude of the coefficients associated with each potential cause of dispersion. Over a period of time it is likely that wage dispersion, technological change, unionisation, female participation, immigration and globalisation have been subject to stochastic trends. Cointegration analysis can be used, to consider whether a linear relationship exists between two or more non-stationary variables, where deviations from this relationship are stationary.

Defining y_t as a vector of the within-group earnings dispersion measure derived from the first step (see equation 2), $v(\hat{\varepsilon})$, technological change (**TC**), globalisation (**G**), female participation (**FP**), immigration (**IM**), and institutional change (**IC**) we have as time series data in each of the four industries:

$$y_t = \left[v(\hat{\varepsilon}), TC, G, FP, IM, IC\right] \qquad \forall_{j,t} \tag{3}$$

where y_t is an unrestricted vector auto regression (VAR) of endogenous variables. Cointegration techniques are employed to see which factor has the largest influence upon within-group earnings dispersion. In particular the cointegrating relationship that we are interested in is where within-group earnings dispersion is defined by:

$$v(\hat{\varepsilon})_t = \phi_1(TC)_t + \phi_2(G)_t + \gamma_1(FP)_t + \gamma_2(IM)_t + \tau(IC)_t + \zeta_t \qquad \forall_{j,t} \tag{4}$$

Table 1: Competing theories capable of explaining within-group earnings dispersion.

Absolute size of coefficients	*Outcome for within-group earnings dispersion*		*Hypothesis not rejected*
If $\|\phi 1\|>\|\phi 2\|$, $\|\phi 1\|>\|\gamma 1\|$, $\phi 1\|>\|\gamma 2\|$, and $\|\phi 1\|>\|\tau\|$	Technological change is the main influence	1	H_1
If $\|\phi 2\|>\|\phi 1\|$, $\|\phi 2\|>\|\gamma 1\|$, $\|\phi 2\|>\|\gamma 2\|$, and $\|\phi 2\|>\|\tau\|$	Globalisation is the main influence	2	
If $\|\gamma 1\|>\|\phi 1\|$, $\|\gamma 1\|>\|\phi 2\|$, $\|\gamma 1\|>\|\gamma 2\|$, and $\|\gamma 1\|>\|\tau\|$	Female participation is the main influence	3	H_2
If $\|\gamma 2\|>\|\phi 1\|$, $\|\gamma 2\|>\|\phi 2\|$, $\|\gamma 2\|>\|\gamma 1\|$, and $\|\gamma 2\|>\|\tau\|$	Immigration is the main influence	4	
If $\|\tau\|>\|\phi 1\|$, $\|\tau\|>\|\phi 2$, $\|\tau\|>\|\gamma 1\|$, and $\|\tau\|>\|\gamma 2\|$	Institutional change is the main influence	5	H_3

Note: Where $\phi 1$, $\phi 2$, $\gamma 1$, $\gamma 2$ and τ are the coefficients on the impact of technology, trade intensity, female participation, immigration and institutional change upon within-group earnings dispersion respectively, from equation 4 in the text.

By comparing the relative sizes of the coefficients this can shed light on the possible determinants of within-group earnings dispersion in each industry, as shown in Table 1.

Provided the data cointegrate then the null hypothesis can be rejected. If the results indicate 1 or 2 from the table (see penultimate column) then this supports hypothesis H1 above, similarly, 3 or 4 supports hypothesis H2, and 5 supports hypothesis H3. However, if none of the industry level data cointegrate with within-group earnings dispersion then the null hypothesis will not be rejected (see above). To test whether the data in equation 3 are cointegrated, and thus give magnitude to the coefficients in equation 4, the Johansen (1988) approach to cointegration is adopted.[4]

The Johansen approach to cointegration is a system estimate based upon maximum likelihood. The data in equation 3 can be reformulated from a VAR into a vector error correction model (VECM) giving

$$\Delta y_t = \Gamma_1 \Delta y_{t-1} + \cdots + \Gamma_{q-1} \Delta y_{t-q-1} + \Pi y_{t-q} + \eta_t \qquad \forall_{j,t} \tag{5}$$

where $\Gamma_d = -(I - A_1 - \cdots A_d)$, d =1...q-1, and $\Pi = -(I - A_1 - \cdots A_q)$. The estimated matrix can be decomposed as $\Pi = \alpha\beta'$ where α and β are (m $\times$ r) matrices and r is the number of cointegrating vectors. The matrix β contains all the cointegrating vectors of the system. The value of r is obtained using sequential likelihood ratio tests, where the presence of a unique cointegrating relationship means that r = 1. The existence of multiple cointegrating relationships implies that $r > 1$ where any linear combination of the columns of β is also a valid representation of the equilibrium relationships. Normalising the inequality measure in the β matrix leads to $\beta = \{1, \phi_1, \phi_2, \gamma_1, \gamma_2, \tau\}\ \forall_{j,t}$ and the cointegraing vector we are interested in as given in equation 4.

4. THE DATA

The first step of the analysis based upon equation 1 requires information on the individual, whilst for the second step more aggregated data at the industrial level are required to gain measures of market forces and institutional changes. Specific factors controlled for in equation 1 are experience, colour, marital status, full/part time employment, highest educational qualification,[5] and regional location. The *GHS* is a continuous survey of cross sections based upon individuals within the sample household. A number of problems were encountered with the data. The definition of earnings changes over time, where post 1979 individuals are asked to report their usual gross earnings. However, no discontinuity is evident over the whole period, and Schmitt (1995) found the data to be consistent with that from the New Earnings Survey. The earnings definition was also revised in 1992 to improve the response to the income section of the *GHS* and so classify more informants by their income. However, the Office for National Statistics (ONS, 1992) reports that the difference in the mean gross weekly income of individuals is not statistically significant pre and post the definition change.[6]

The main problem encountered with the *GHS* data was the change in the Standard Industrial Classification (SIC) in 1980, where the industry codes were reduced from twenty four groups to ten in the *GHS*. However, to find out which industries are consistent over the period the ratio of the industry sample size to total sample size based upon the matched categories pre and post the SIC break was found and then the percentage change from the previous year calculated - for details see Taylor (1999). Those industries where the absolute percentage change in the break year 1980 to 1981 was greater in magnitude than other years were excluded. The resultant industries found to be consistent over the entire period based upon the above methodology were: Manufacturing SIC 3, Other Manufacturing SIC 4, Construction SIC 5, and Transport & Communication SIC 7.

More aggregated data, used in the second stage of the analysis, are required at the industry level in order to attempt to find the possible causal factors of within-group earnings dispersion. Such explanations come in the form of market forces (demand and supply factors) and institutional change, where it is required to find some proxy for each. On the demand side technological shocks are proxied by research and development intensity for each industry. This is defined as research and development expenditure as a proportion of value added, using data from the OECD ANBERD data base and OECD STAN data base respectively - with all expenditure data deflated to 1973 prices. Globalisation in the traded sector (SIC 3 and SIC 4) was proxied by trade intensity, defined as import expenditure as a proportion of value added. The source of the trade expenditure data was also the OECD STAN data base - again all expenditure data was deflated to 1973 prices. For the supply side, immigration and female participation by industry was derived from the *GHS*, and was calculated as those individuals born outside the United Kingdom (female) who were in employment (defined as working more than one hour per week) as a ratio to total industry employment size.

A measure of institutional change proved to be a relatively more difficult task than at first sight. The preferred measure to be used would have been trade union density or membership. Unfortunately the figures are only available consistently at an aggregate level from the Department of Employment. Previous researchers namely Bain and Price (1983) have constructed one digit industry level trade union membership and density, but only up until 1979. Thereafter the source they use, namely the *Labour Force Survey*, does not collect union data at the industry level for each proceeding year. Thus in an attempt to proxy institutional change the number of workers involved in strikes for each industry based upon International Standard Industrial Classification codings was used, available from the International Labour Organisation. Strike action represents one form of bargaining power, where a threat to strike is credible if the firm cannot replace its workforce easily. Consequently, the extent of unionisation and the ease of substitutability between union and non-union members are of importance. The analysis of the second stage uses strikes to proxy for institutional change as it follows the trend in union membership - at the aggregate level a correlation of 0.9 is observed. This is consistent with previous findings (Machin, 1997).

Summary statistics of the industry level data are shown in Table 2. Clearly R&D intensity was highest in manufacturing averaging 9 per cent over the period from a low of 6.3 per cent in 1975 to 10.2 per cent by 1993. Trade intensity is higher in manufacturing than other manufacturing by around 15 percentage points each year, although it displays the same amount of variability in both industries. On the supply side both female participation and immigration show approximately the same variability across time for each industry and for immigration each industry experiences less than 10 per cent on average. Female participation is largest in other manufacturing and transport & communication where it has increased from around 20per cent in 1973 to over 25 per cent by 1995.

Looking at industrial disputes, through the number of workers involved in strikes, shows manufacturing and transport & communication with the largest number of workers involved. However, over time the number of people involved in strike activity has fallen by 1995 - in manufacturing to just 400, in construction to 2000 and in transport & communication to 200. The only industry not to experience such trends was other manufacturing where strike activity in 1995 was 32,800, noticeably above the mean.

Table 2: Industry level sample statistics

	Mean	*Standard deviation*
Manufacturing		
Within-group dispersion	0.324	0.06
R&D intensity	9%	1.27%
Strikes	460,426	505,541
Trade intensity	31%	10.4%
Female participation	21%	2.52%
Immigration	7%	1.28%
Other Manufacturing		
Within-group dispersion	0.363	0.05
R&D intensity	1%	0.14%
Strikes	13,374	13,918
Trade intensity	17%	9.8%
Female participation	40%	4.32%
Immigration	7%	1.09%
Construction		
Within-group dispersion	0.341	0.08
R&D intensity	1%	0.31%
Strikes	16,089	14,204
Female participation	10%	3.17%
Immigration	5%	1.28%
Transport and communication		
Within-group dispersion	0.347	0.08
R&D intensity	6%	1.60%
Strikes	117,352	112,923
Female participation	21%	2.55%
Immigration	8%	1.28%

5. Empirical results

The first part of this section gives the results of decomposing earnings dispersion into between and within-group components, considering the trend in earnings dispersion for each industry, and whether each industry experienced a rise in demand for higher skilled workers. The second part considers what may have caused within-group earnings dispersion in each industry.

Within-group and between-group components

For each industry the log of weekly wages was regressed against the explanatory variables given in equation 1 above,[7] where the estimation technique used was generalised least squares in order to control for heteroscedasticity (which if present would be likely to inflate the measure of within-group earnings dispersion upwards). Each industry specific cross sectional equation was tested and found to be robust to chosen functional form, parameter stability, the influence of outliers and the role of changing hours over time upon dispersion (for details see Taylor, 1999).

Table 3: Pairwise correlations of within-group earnings dispersion across industries.

Sample 1973-95	*SIC 3*	*SIC 4*	*SIC 5*	*SIC 7*
SIC 3	1.000	0.9178	0.9132	0.8056
SIC 4		1.0000	0.8518	0.7754
SIC 5			1.0000	0.7859
SIC 7				1.0000
Sample 1973-77	*SIC 3*	*SIC 4*	*SIC 5*	*SIC 7*
SIC 3	1.0000	0.5133	0.8624	0.0046
SIC 4		1.0000	0.6793	0.0155
SIC 5			1.0000	0.1957
SIC 7				1.0000
Sample 1978-83	*SIC 3*	*SIC 4*	*SIC 5*	*SIC 7*
SIC 3	1.0000	0.3959	0.6814	0.5791
SIC 4		1.0000	0.1149	0.0914
SIC 5			1.0000	0.7759
SIC 7				1.0000
Sample 1984-89	*SIC 3*	*SIC 4*	*SIC 5*	*SIC 7*
SIC 3	1.0000	0.4600	0.0094	0.1861
SIC 4		1.0000	0.0316	0.6378
SIC 5			1.0000	0.1437
SIC 7				1.0000
Sample 1990-95	*SIC 3*	*SIC 4*	*SIC 5*	*SIC 7*
SIC 3	1.0000	0.8299	0.8693	0.5458
SIC 4		1.0000	0.8223	0.5404
SIC 5			1.0000	0.4429
SIC 7				1.0000

SIC3=manufacturing, *SIC4*=other manufacturing, *SIC5*=construction, *SIC7*=transport and communication

Figures 1 to 4, below, show the trend in overall earnings dispersion, and the contribution of between-group and within-group dispersion over time, for each of the four industries. In each of the four industries within-group earnings dispersion dominates between-group earnings dispersion . Noticeably each industry experienced differing trends in earnings dispersion, for example in Manufacturing and Construction within-group earnings dispersion remained roughly constant from 1973 to 1983, whilst in the other two industries it fluctuated.

Figure 1: Earnings dispersion in manufacturing.

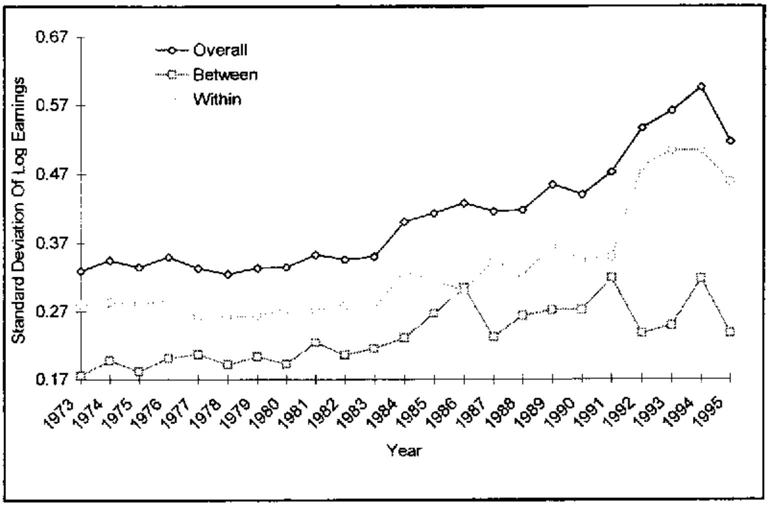


Figure 2: Earnings dispersion in other manufacturing.

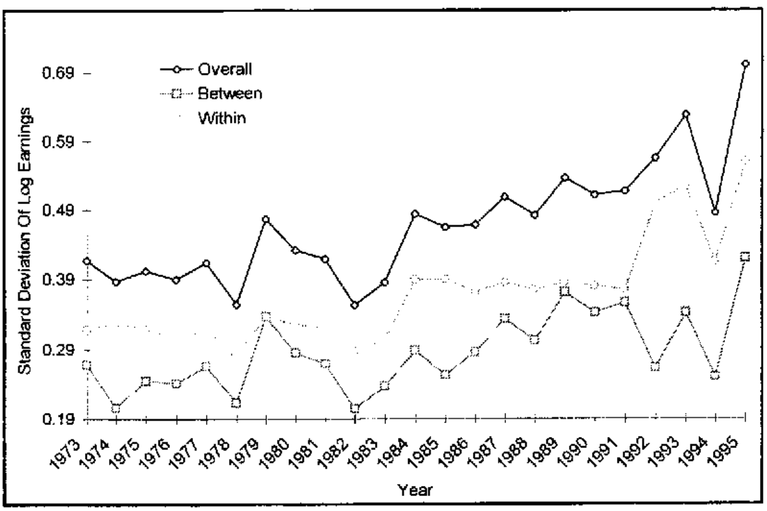


Figure 3 Earnings dispersion in construction

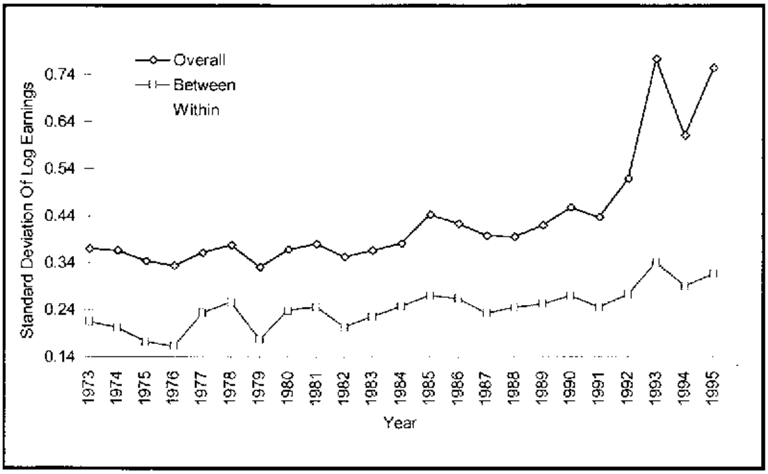
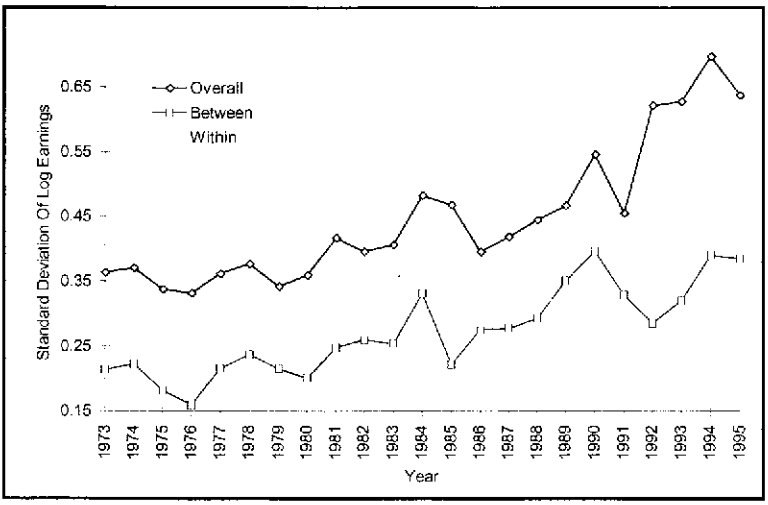


Figure 4 Earnings dispersion in transport and communication



This is made clearer in table 3, above, where pairwise correlations between within-group earnings dispersion in each industry are shown. The first part of table 3 shows the correlation over the entire sample period - where clearly the trend of within-group earnings dispersion in manufacturing is highly correlated with other manufacturing. Splitting the sample into four periods - 1973 to 1977, 1978 to 1983, 1984 to 1989 and 1990 to 1995 - shows that the data trend in different ways across industries. If the measure of earnings dispersion has a similar trend across industries then in each time period in table 3 the pairwise correlation should be high - however this is not the case. For example, in manufacturing the correlations with other industries fall over each sub-period until 1990 to 1995. This is also generally true for construction and other manufacturing. This makes a strong argument for analysing industry level trends, rather than just manufacturing or economy level earnings dispersion. Clearly, in each industry overall earnings dispersion rose and it is evident that this was due to increasing within-group earnings dispersion. This pattern of within-group dispersion dominating between-group dispersion is consistent with previous research findings and is evidence that the demand for relatively higher skilled workers has increased in each industry (Juhn *et al.*, 1993 for the United States; and for the United Kingdom, Schimtt, 1995 and Machin, 1996a).

Having briefly discussed the results of the first stage of the empirical methodology, the following analysis attempts to provide an explanation for any remaining earnings dispersion, that is within-group earnings dispersion. Not only has the demand for observable skills risen in the economy (Machin, 1998) and at the industry level it appears that the demand has also risen within certain groups. What remains to be seen is what has caused these demand changes influencing within-group earnings dispersion.

Results from a time series analysis of within-group earnings dispersion

Having discussed the results from the first step of the empirical process, and found that each industry experienced different trends in earnings dispersion, the following looks at the results from the second stage of the empirical approach. The second part of the empirical analysis is based upon a time series approach, where initially the order of integration of the data is checked. Because of the problems associated with non-stationary data the following tests to see if the measure of within-group earnings dispersion, market forces and institutional change data exhibit unit roots. The results of stationarity tests based upon the ADF procedure are shown in table 4, where the first column tests the null hypothesis that the data contains a unit root by considering the data in levels. The null hypothesis cannot be rejected for any of the data and so is presumed to be non-stationary. Due to the problems associated with the ADF test (see above) the KPSS test was also applied to the data. Unlike with the ADF test, the null hypothesis is of stationary data and so will not be rejected unless there is strong evidence in contrast to the usual case of non-stationary data. The results shown in final column of table 4, above, reject the null hypothesis of stationary data in each instance, usually at the 1 per cent or 5 per cent level. Thus the KPSS test for unit roots rejects the null hypothesis of stationary data, supporting the standard ADF test.

Having accepted that the data are non-stationary, the following employs the Johansen approach introduced in section 3. Cointegration rank is tested by the trace statistic and the maximal eignenvalue test, each based upon unrestricted intercepts and restricted trends in the underlying VAR, as shown in table 5. Both statistics reject the null hypothesis of no cointegration at either the 1per cent or 5per cent level, and show a rank of unity across industries - in other words only a single cointegrating relationship is found. Each variable is tested for autocorrelation (AR), heteroscedasticity (HET), normality (NORM) and auto-regressive conditional heteroscedasticity (ARCH). Also, tests of autocorrelation and normality in its multi-variate form were carried out by industry. The results, not reported, found that at either the 1per cent or 5 per cent level of significance all the diagnostics could be passed in each industry.[9]

Table 4: Unit root tests

	ADF test	KPSS test
Manufacturing		
Within-group earnings dispersion	2.54	0.3031***
R&D intensity	1.16	0.1376*
Strikes	1.49	0.1435*
Trade intensity	1.45	0.8682***
Female participation	1.79	0.3015***
Immigration	2.51	0.2992***
Other manufacturing		
Within-group earnings dispersion	2.68	0.3032***
R&D intensity	2.25	0.1231*
Strikes	2.56	0.1823**
Trade intensity	2.29	0.1539**
Female participation	2.55	0.3022***
Immigration	3.77	0.3004***
Construction		
Within-group earnings dispersion	1.15	0.3028**
R&D intensity	1.22	0.1223*
Strikes	3.14	0.2307***
Female participation	2.75	0.2962***
Immigration	3.77	0.2957***
Transport & Communication		
Within-group dispersion	1.46	0.3020***
R&D intensity	2.24	0.1577**
Strikes	1.44	0.1393***
Female participation	2.79	0.3013***
Immigration	2.87	0.2990***

Notes: *** Significant at the: 1% level, ** 5% level, and * 10% level. Perron tests performed upon the data (not shown) rejected the hypothesis of structural breaks in favour of unit roots. Even if structural breaks are present in the marginal process for a particular variable Campos *et al.* (1996) find that test statistics from multivariate cointegration approaches are unaffected by breaks.

Table 6 below shows the size of the impact and its direction for each explanatory factor across industries, with t-ratios in parenthesis based upon Newey-West (1987) standard errors. The magnitudes of the coefficients are presented in the form of equation 4 and are ranked from 1, the largest impact upon within-group earnings dispersion, downwards. In both manufacturing and other manufacturing the main two culprits for increasing dispersion, were trade and technology.

These results are consistent across different methodologies for testing cointegration (see endnote 3) and find that trade has the largest impact in explaining within-group earnings dispersion in manufacturing and other manufacturing sectors.

Table 5: Tests of cointegrating rank adjusted for sample size

	Rank	*Trace statistic*	*Maximal eigenvalue statistic*
Manufacturing	r=0	46.84**	98.76*
	r<=1	27.5	51.92
	r<=2	11.64	24.42
Other Manufacturing	r=0	41.45*	96.95*
	r<=1	27.36	55.5
	r<=2	14.01	28.14
Construction	r=0	40.12*	97.9**
	r<=1	24.04	52.01
	r<=2	15.31	33.73
Transport & Communication	r=0	40.77*	88.29**
	r<=1	20.7	47.53
	r<=2	14.93	26.83

Note: *** Significant at the: 1% level, ** 5% level, based upon distributions from Osterwald-Lenum (1992).

What is perhaps surprising is the role played by female participation in manufacturing and transport & communication, and immigration in all industries excluding manufacturing. This can be interpreted as female participation and immigration having acted as substitutes to lower skilled labour and hence increases in their supply over time has resulted in falling relative demand for lower skilled labour, as their signs suggest. This would then be consistent with having caused a demand shock and the fact that female participation has increased the supply of skilled labour (Harkness, 1996). Similarly, the argument that immigrants are becoming more skilled over time relative to native males, and so are possible substitutes to lower skilled native males, is more plausible and consistent with the findings of Bell (1997) who found that the gap between the educational attainment of immigrants and native males in the UK has grown in favour of immigrants over the period 1973 to 1992, that is successive cohorts of immigrants are more educated than native UK males.

Also surprising is that the impact of institutional change is ranked last in manufacturing and other manufacturing and at best only third in construction. Institutional change also has a negative coefficient in manufacturing suggesting that changes to labour market institutions have compressed the wage distribution in this industry. This raises the issue of whether strikes are a good measure for the decline in institutional arrangements in particular collective bargaining - this is discussed at length in Taylor (1999) where it is argued that the strike variable does a fairly good job of measuring the decline in collective bargaining. This leads to the question that if the trend in our measure of institutional change follows the same trend as other measures then why is its impact so small in each industry? One possible response is that once one looks at the 1990s the strike data doesn't capture the trend in unionisation that well due to the very large fall in strike activity, see Machin (2000). Previous findings have indicated a large role for institutional change (Gregg and Machin, 1994). One possibility is that previous work has typically only tested the impact of institutional change upon dispersion, without including other potential factors such as technological change etc., (for example, Leslie and Pu, 1995, 1996) and so consequently this analysis has an advantage in that other factors are controlled for - with the results implying other factors have a larger role to play. Also previous work has typically only taken a snapshot for particular years rather than a long time series, the exceptions for the UK are Leslie and Pu (1995, 1996).

Table 6: Cointegration results

	Coefficient	*T-ratios*	*Rank*
Manufacturing			
R&D intensity	0.4516	3.225	2
Strikes	-0.1679	2.331	5
Trade intensity	0.6138	8.184	1
Female participation	0.2993	4.129	3
Immigration	0.2909	6.402	4
Other Manufacturing			
R&D intensity	1.4296	5.546	2
Strikes	0.1569	7.411	5
Trade intensity	3.0454	9.427	1
Female participation	0.4399	2.225	4
Immigration	0.9578	7.041	3
Construction			
R&D intensity	0.2786	4.413	1
Strikes	0.1317	12.430	3
Female participation	-0.0468	0.859	4
Immigration	0.2517	6.985	2
Transport & Communication			
R&D intensity	0.0406	2.919	3
Strikes	0.0076	2.347	4
Female participation	0.0796	2.860	1
Immigration	0.0709	3.295	2

Note: Rank 1=greatest impact, 4/5=smallest impact. T - ratios are shown in parenthesis based upon Newey-West standard errors.