Chapter 9

Male earnings 1977-1990: fixed-effects and varying coefficients

This chapter documents the results of estimating straightforward fixed-effect (FE) and cross-section wage equations for males using the NESPD from 1977 to 1990. These results are notable in a number of ways. First, they provide a complete set of cross-sectional estimates for the period 1977-1990 allowing the coefficients to vary over time. Thus it is possible to study the unrestricted evolution of the coefficients over fourteen years. More importantly, the fixed-effects estimator is used to generate panel estimates. Although some cross-sectional estimates have been made, these are the first true panel estimates from the disaggregated NESPD. The evidence to be presented here suggests that allowing for individual heterogeneity has a significant impact on the results.

Of some interest in their own right, these results are also used to compare the effect of differing estimators on the results obtained. This comparison takes two forms. Firstly, the results of the fixed-effects specification are contrasted with cross-section (CS) estimates (the models of sections 5.2 and 5.1). Secondly, the ability to let coefficients vary over time is used to consider the evidence for parametric stability in the UK labour market over the period.

9.1Econometric issues

9.1.1Functional form

The framework for estimation is the Mincer-type reduced form specifying log earnings as a function of "human capital" and other control variates; more specifically, for the FE estimator the unrestricted equation (5.45) is the basic specification:

$$w_{it} = x_{it}\beta_t + \lambda_t + \alpha_i + u_{it} \qquad \textbf{(9.1)}$$

This model has time-varying coefficients and fixed individual effects, and is referred to as the

TVFE estimator. The cross-section model ignoring the individual-specific term α_i still has time varying coefficients, and henceforward is referred to as the TVCS model[1].

The nature of the reduced form specification means that the presence of the controls can be given a wide variety of interpretations, reflecting various theories of wage determination. For example, regional dummies might reflect compensating differentials or differences in the pressure of demand in local labour markets due to geographic immobility of the labour force. Because the reduced form is consistent with many structural models of the labour market it may be unable to discriminate between them. Therefore the results reported here are not interpreted with respect to any one particular hypothesis.

Given the size of the NES, the asymptotic properties of the estimated coefficients are an important consideration. As the TVFE estimator is fundamentally OLS, implicit assumptions of the results in this section are that

$$E(x_i' u_i) = 0 \qquad \operatorname*{plim}_{n \to \infty} x_i' u_i = 0 \tag{9.2}$$

where $\mathbf{x}_i \equiv [\mathbf{x}_{i1}'..\mathbf{x}_{iT}']'$ and $\mathbf{u}_i \equiv [u_{i1}..u_{iT}]'$. The validity of these assumptions is debatable. It could be argued that all the variables in the NES are potentially endogenous[2]. However, the effect of this endogeneity is unknown. Therefore, in this particular study, the null hypothesis is that sufficient exogenous variables have been included to avoid omitted variable bias so that (with the allowance for individual heterogeneity) the assumptions in (9.2) hold. Although further work may refute this hypothesis, it seems a reasonable starting point[3].

[1] The TVFE and TVCS specifications are the "unrestricted" equations (5.45) and (5.1), respectively. The terms TVFE and TVCS are used in future to avoid confusion over the use of "unrestricted" in describing the models.

[2] For example, self-selection (ie labour supply) by the NES sample could involve wages and hours offered, overtime rates, occupation, region, industry, union status and the predilection for full-time work. Self-selection by employers in the NES (labour demand) could involve wages and hours desired, occupation, industry, union status, public/private sector, and so on. These two lists contain all the NES variable fields.

[3] In the following chapter the issue of endogeneity is considered in more detail.

The TVCS model estimates (9.1) as T separate cross-sectional analyses:

$$w_{it} = x_{it}\beta_t + \lambda_t + \eta_{it} \qquad \eta_{it} \equiv \alpha_i + u_{it} \tag{9.3}$$

Clearly, if α_i is non-zero, η_{it} will appear to be serially correlated and OLS estimates of (9.3) will be inefficient. More importantly, CS estimates of (9.3) involve three extra assumptions:

$$E(\alpha_i) = 0 \quad E(x_{it}\alpha_i) = 0 \quad \underset{n\to\infty}{plim}\, x_{it}\alpha_i = 0 \tag{9.4}$$

The first assumption is not important as the individual intercept can be split into a mean common for individuals and deviations from that mean. Thus a non-zero mean for the individual-specific effects is simply subsumed into the time-intercept for those appearing in a particular period. The second and third assumptions are more important, implying additional restrictions on the regressors. Any potential correlation between the job characteristics and the invariant characteristics of the individual must be discounted (see section 2.2).

In the context of the NES, independence of the characteristics of the individual and the job is a large assumption. A number of authors (for example, Chamberlain (1985); Hartog and Oosterbeek (1993); Jakubson (1991); Killingsworth (1986); Rees and Shah (1992)) have noted that individual heterogeneity may influence occupation, sector, location, and so on. A related argument which has seen more attention in the literature on female earnings is that significant determinants of labour market experience are pre-entry decisions and the initial job taken[4]. If the choice of first job is non-random and significant in determining future employment, this may lead to an additional selection bias in CS models[5]. Most importantly, a premarket factor influencing future employment is likely to be education, which the NES omits. Given that formal (certified) education is usually completed before employment, and is thus a time-invariant individual characteristic as far as employment history is concerned, assumption (9.4) should be treated with caution.

[4] Elliott(1991) pp404-407 discusses some aspects. Empirical analyses on pre-entry influences include Dolton and Mavromaras(1994) on expectations of career prospects; and Vella(1994) on sociological attitudes.

[5] The reason this can lead to a bias in CS and not FE models is that the initial job choice may manifest itself as a "one-off" influence which is constant throughout an individual's working life; in other words, a fixed effect. See Ridder (1990) or Verbeek and Nijman (1992a).

9.1.2Hourly versus weekly wages

Both hourly and weekly wages have been used as dependent variables in labour market studies. It may be argued that weekly pay is a better indicator of the slope of the budget constraint faced by workers given that standard hours of work are usually fixed by the employer rather than through direct negotiation. If the marginal value of leisure hours is relatively constant over the week and there is little flexibility in the relationship between working time and wages (for example, there is little or no overtime premium), then demand and supply functions based on the weekly or annual earnings may be appropriate. This measure is most likely to be the case where working time is not a significant determinant of the labour supply decision (for example, for salaried employees or non-manual employees for whom overtime is not generally available).

However, the dependent variable used in this study is hourly wage. The main argument is that this is a better indicator of marginal benefits and costs, as that the wage rate and the number of hours worked form two separate (if not independent) choice criteria[6]. This allows for the joint nature of the income/effort decision by employees and the employee/working time decision by employers. Where the marginal value of leisure varies (for example, evening working requires more compensation than Saturday jobs) an hourly rate more accurately measures the leisure/income tradeoff at the margin; and where a range of working practices and incentives is available to the employer the hourly rate arguably represents the marginal cost of labour more truly.

Put more formally, in an individual-level study it is reasonable to assume that the individual maximises a utility function which contains both hours worked and leisure:

[6] A number of authors have argued that the hours/wage decision is made simultaneously (see Killingsworth (1983), MaCurdy(1985) or Stern (1986) for surveys) which will not be considered here. The key point is that the wage rate determines participation levels rather than total income.

$$U = U(f(h_{max} - h), wh) \quad U_w > 0, \ f_h \leq 0 \qquad \textbf{(9.5)}$$

where h_{max} is maximum hours available for work, h is hours worked, w is the hourly wage rate and the function f(...) captures the utility of leisure time. In this context, using weekly wage (wh) is only appropriate if f_h=0 or if h is not a choice variable for individuals. In the latter case, which may be appropriate for non-manual workers, hourly and weekly wages only differ by the scaling factor h and weekly wages will capture the marginal value of work accurately. If, however, workers have some control over h, then only the hourly wage is appropriate.

However, the choice of *the* wage rate is problematic when all hours are not paid at the same rate. For example, if overtime and weekend premia are all available to the employee, then the marginal wage rate may differ dramatically from the average wage rate which is typically reported. When the wage function is significantly non-linear, then a simple mean wage averaged over all hours will not reflect the employee's labour supply function (Brown, Levin, Rosa, Ruffell and Ulph (1986)). Similarly, if untimed payments are made (such as production bonuses) then the allocation of these bonuses to wages may be arbitrary and unjustified. Thus the labour supply decision is likely to involve a range of possible "wages" at both weekly and hourly rates. Moreover, the difficulties of measuring hours of work for non-manual employees means that often the hours of non-manual workers are concentrated around a standard hourly week and do not reflect the actual hours worked[7].

The choice is further complicated in the NES as the reported hourly wages are the <u>actual</u> hourly wage experienced during the survey week, defined as wages for that week divided by the number of hours actually worked. In other words, only weekly wages and hours worked are known. Information on overtime hours and bonus payments is available, but not on whether such payments are typical or atypical; thus, it is difficult to say whether the

[7] Atkinson, Micklewright and Stern (1982) compare employee perceptions of hours worked (from the FES) with employer perceptions from the NES. While for manual workers the two are similar, for non-manual workers employees believe they work much longer hours than their employers think they do.

wage received represents the "normal" distribution of wage offers and therefore gives a "fair" view of the remuneration options open to the individual.

Despite these difficulties, the wage rate used here is the natural logarithm of actual observed hourly wages excluding overtime payments, adjusted for RPI. Given the nature of this study, to provide results from a new estimator on a familiar dataset, it takes the approach of the bulk of the literature (see Killingsworth (1983, especially tables 3.1, 4.1, and 5.1) for a comprehensive survey of earlier results).

However, it may be noted that using the weekly wage as the dependent variable makes little qualitative difference to most estimates. Bell and Hart (1995) study the question of hourly versus weekly wages and basic versus total compensation in some detail, and report that the choice of measure makes relatively little practical difference. This was in respect of one variable only, the "union markup", and so may not hold for other variables, but a cursory comparison of the TVCS study of Andrews, Bell and Ritchie (1993) and the TVCS results presented here suggests that the only notable difference between using weekly and hourly earnings is to be found in the sectoral coefficients (see section 9.2.9).

9.1.3Attrition and missing data

A serious, but largely unrecognised problem with the NES is the large amount of missing data. Missing data has two effects: it reduces the precision of estimates; and, if correlated with the variables of interest, it can invalidate the estimation results. This problem is, of course, not unique to the NES or panels, but the nature of the dataset makes it difficult to counter.

A descriptive analysis of the missing data problem in the NES has been attempted in Bell and Ritchie (1994). The general conclusion of this work is that the likelihood of selection bias is large; the probability of individuals and observations being included in the dataset appears to be correlated with almost all the variables in the dataset. This is not surprising given the range of characteristics covered by the NES and the potential for complex variable relationships, but it is a source of concern. However, the *magnitude* of the effect of missing data is much harder to identify, especially as the size of any effect is specific to the particular equation being estimated.

As noted in chapters two and four, the construction of models of attrition for panels is complex in practice and requires strong assumptions. The standard econometric approach to dealing with non-random attrition in a two-period model was developed by Hausman and Wise (1979). Absence from, or presence in, the panel is modelled in a separate probit equation and a Mills ratio for each individual derived from this equation is included in the earnings relationship. Unfortunately, the Hausman and Wise solution is derived from the simplest possible panel model; in a multi-period model this procedure is computationally prohibitive, even if the size of the NES did not preclude non-linear estimation. It is also of doubtful value if a dynamic specification of attrition is desired.

Following a suggestion of Verbeek and Nijman (1992a), the effect of absence (from the panel and the workforce) is linearly approximated by including as additional regressors variables which are related to the probability of attrition. This is a simple approach in the present context, since, for example, it is straightforward to calculate whether an individual was present in the previous year, how many previous years they had been present and so on.

This is not as ad hoc as it seems. There is a close relationship between the switching-regression adjustments for selection bias commonly used and the linear instrumental variables approach (Vella and Verbeek (1993)), and the proxy-variables method is similar to instrumenting selection dummies with the approximations. This result has also been noted by Lanot and Walker (1993a), who use IV as one of several methods for correcting union membership selection bias. Their results indicate that results may not be sensitive to the particular correction method used. This is not an unexpected result, given that the switching regression approach merely requires *consistent* estimates of the selectivity term, not efficient ones.

Necessary assumptions for this approach to be fully effective are that a linear term is adequate to represent the form of attrition bias; that the variables chosen are sufficiently good proxies for attrition; and that the variables themselves do not lead to further bias or inconsistency in the estimates. These are strong assumptions, but given the nature of the estimates here presented, and the scope for selection bias, the net effect of adding proxy variables is likely to improve the specification[8]. A side-effect of introducing these proxy variables is that significance tests can give an indication of whether selection bias is important, although the approximate nature of the variables makes these tests of debatable worth (it should also be noted that any estimates would describe observation in the panel rather than employment patterns). These variables also need to be treated with care in cross-sectional studies, which may confuse them with individual heterogeneity.

9.2Fixed-effects and cross-section results

In addition to the basic TVFE and TVCS models, "pooled" and "restricted" models (to use the terminology of chapter five) were also estimated for both the FE and CS models. This gives six basic specifications.

$$\begin{array}{lll} (a) & w_{it} = \lambda_t + x_{it}\beta_t + \alpha_i + u_{it} & TVFE \\ (b) & w_{it} = \lambda + x_{it}\beta + \alpha_i + u_{it} & FE\ pooled \\ (c) & w_{it} = \lambda_t + x_{it}\beta + \alpha_i + u_{it} & FE\ restricted \\ (d) & w_{it} = \lambda_t + x_{it}\beta_t + u_{it} & TVCS \\ (e) & w_{it} = \lambda + x_{it}\beta + u_{it} & CS\ pooled \\ (f) & w_{it} = \lambda_t + x_{it}\beta + u_{it} & CS\ restricted \end{array} \qquad \textbf{(9.6)}$$

The pooled and restricted models are those discussed in chapter 5: (5.61) and (5.74) for the FE specification, and (5.20) and (5.27) for the CS. This section presents results of estimation on (9.6a) and (9.6d); the next section uses the restrictions to consider the question of parametric

8 Further work to test the validity of the proxy-variables approach is being carried out.

stability.

The explanatory variables used were: the attrition variables (AVs); occupation (CODOT grouping); region; industry; age; private sector; coverage by collective agreement; coverage by Wages Council; and job held for less than one year. The last four of these are single dummy variables; the others, save the AVs, are all categorical variables. Codes are given in an appendix to the chapter.

Three AVs were employed:

InLast1 iff in the panel last period; 0 otherwise

YrsInNumber of years in the panel to date

CurrStayLength of current continuous run in the panel

Regressions were run on the period 1977-1990 (1975-1990 to calculate the AVs). Because of the large amount of output (797 coefficients are estimated; fifty-seven variables over fourteen years less one time dummy), only a sample of output is given in table A9.2 in the appendix. The sample year (1984) is half-way through the period under review. Full results are available on request.

9.2.1TVFE versus TVCS specifications

A first consideration is whether the TVFE and TVCS models are significantly different, for the TVFE model is computationally more involved than the TVCS model. As the TVCS is nested within the TVFE model, F-statistics can be constructed readily to test the null hypothesis that the TVCS restrictions are justified. Summing the fourteen TVCS RSSs and comparing with the TVFE RSS gives a test F-statistic of 12.22 with {194782, 907324} degrees of freedom, rejecting the null hypothesis (see table A9.1 for summary statistics). This is a reasonable result for, as noted in section 9.1.1, individual characteristics are likely to be related to both earnings and job characteristics.

Of more interest is how close the TVCS results are to the TVFE results, as the former are much easier to estimate than the latter. The estimation results to be presented here indicate that the *qualitative* features of the two models are similar, but the *scale* differs - dramatically in some cases. As the TVCS and TVFE results are generally consistent with theory and other research, the qualitative similarity is not surprising. However, the scale differences suggest that cross-section results for the NES are likely to be significantly biased.

TVCS results are biased away from zero in most cases: for all but two variables (agreement coverage and the attrition variable YrsIn), the FE model reduces the absolute coefficient values. In other words, after allowing for individual-specific heterogeneity, the variation in returns to a characteristic are much smaller. This supports the view of section 9.1.1 that the explanatory power of some variables in cross-sections is due to the correlation between earnings, unmeasured individual heterogeneity and job characteristics.

One other general comment on the TVFE/TVCS results is that the TVFE estimates tend to be much smoother over time. This is to be expected given the greater efficiency of the TVFE estimator and the susceptibility of the TVCS model to outliers in particular years.

The specific results are now considered in more detail.

9.2.2Means and constants

Figure 9.1 displays the constant terms calculated by the two models for all three specifications in (9.6), along with the mean of the dependent variable ln(hourly wage). The rise of mean earnings over time reflects the trend growth in wages, as the earnings figure is adjusted for RPI rather than a general wage index. However, note the decline in the earnings of the representative individual up to 1979.

The FE results give a smoother path for the representative individual over time, as the CS mean is more affected by the changing composition of the panel. The lower value for the conditional mean wage in the TVFE model indicates that the average individual-specific intercept is positive (that is, $E(\alpha_i)\neq 0$ in (9.4)); in other words, the unmeasurable component of wages is larger than in the CS model and the measurable effects smaller.

9.2.3Region (reg)

The reference category is Greater London. Figure 9.2 shows that the return to working in London as opposed to other regions increased steadily over the period. The largest differential was with northern England and Scotland, whilst those living in the south-east and East Anglia saw the smallest drop in their relative earnings. The CS model over-estimates the coefficients (compared to the FE model) by around 50% in the early years when inter-regional differences are relatively small.

Layard and Nickell (1987) argue that differences in inter-regional wage pressure lessened somewhat throughout this period in terms of regional unemployment-to-vacancies ratios, which should have led to a fall in the regional variation in wages, ceteris paribus. The apparent contradiction of figure 9.2 is because the Layard and Nickell do not take account of the regional characteristics, whereas the regression results are conditioned on industry, occupation, and sector and imply a "pure" regional effect which supports anecdotal evidence that the South prospered relative to the North over the 1980s. If the country is crudely characterised as a manual, manufacturing, low demand North and a non-manual, predominantly services-based high demand South, then these studies are consistent (especially as Layard and Nickell note a small rise in "industrial mismatch")[9].

---

[9] The Layard and Nickell measure of wage pressure also only takes account of total unemployment, whereas many authors (including Layard and Nickell; note 21, p175) have argued that both the number and type of the unemployed affects wages; see Ham (1986) and Theodossiou (1992) for example.

9.2.4Industry (div)

Figure 9.3 depicts the improvement in earnings in all industries relative to farming and fishing (FF) over the fourteen years. Although earnings in this industry start off around the average level for all industries, by 1990 FF has the lowest remuneration. One difference between the FE and CS models is that the former puts FF at the bottom of the earnings heap almost immediately, whilst for the latter it is not until 1982 that FF really moves into the low-pay bracket.

More distinctive is the result for Banking, Finance and Insurance (BFI: group 8). The cross-section moves the increment to pay from 0.05 to 0.25, a relative move in line with other industries. However, the FE estimates show the increment moving from the bottom of the range (at -0.14) to almost the top (at +0.17).

Thus, once individual differences are allowed for, the BFI group has made a very large advance in relative earnings since 1977. The rise throughout the 1980s can be put down to increasing financial deregulation and the boom in financial services. More surprising is the relatively poor initial state. This may be due partly to a "cohort effect" - younger employees with high earnings and high earnings growth push down the relative wage of older cohorts. An alternative is that branch banking employees have relatively low wages and wage growth compared to those with similar qualifications in similar jobs; before the growth in financial services these constituted the bulk of employees in this sector. However, this is largely speculation without more detailed information on this group.

It may also be noted that, apart from the two groups mentioned, the other industries maintain much the same relative position over time. This result has also been reported in the US (Helwege(1992)), where there is a wide literature discussing the "efficiency wage" view of inter-industry differentials (see Krueger and Summers (1988), for example).

9.2.5Occupation (kos)

The default occupational grouping is "clerical and related" - the basic non-manual group. Figure 9.4 gives the relative positions of the manual groups over this period. The results for both models are similar in shape, but the cross-section shows a much larger absolute divergence and a relative upward shift in the position of the non-manual worker. Allowing for individual heterogeneity, the non-manual worker is comparable with the average manual worker, although his relative position has steadily improved since 1977.

The change in the relative position of security and protection (group 254) stands out. Although the cause of this change is not clear, both TVFE and TVCS estimates show a steady improvement until 1985, then a falling off for the rest of the period. As individual-specific fixed effects do not affect this shape, it may be that the change in the relative position of this group is due to demand rather than supply factors.

Robinson (1994) notes that "low-paid" (including clerical) jobs have declined in importance for men since the war in terms of numbers employed; employment growth has been in "high-paid" (professional and managerial) occupations. Figure 9.5 shows that, compared with other non-manual workers, clerical workers have become steadily worse off during the 1980s. Thus professional and managerial workers have not only improved their relative earnings but also their share of employment, which suggests that the increased returns to this group are due to increased demand. This is less clear in the CS figures because the scale is much larger. One reason for the huge difference in the size of the returns to occupations may be that non-manual occupations rely much more on "unmeasurable" characteristics: personality, motivation, talent, ability, and so on. If these characteristics stay roughly constant over the period of the survey, this could explain the disparity between the TVFE and TVCS findings.

A second significant factor in the differences between the estimates may be education. Greenhalgh (1980) notes the relationship between occupations and levels of education. For the reasons noted in section 9.1.1, education is likely to produce an individual specific element correlated with occupation which would be transformed out by the TVFE model. Thus the relatively poor performance of the TVCS model may be due to omitted variable bias.

Assuming that education and occupation are not significantly correlated (and so both the TVFE and TVFE are consistent and unbiased), this raises the issue as to which model gives the "better" result. In section 9.1.1 it was noted that the assumed mean of these effects is zero - implying that a non-zero mean is subsumed into the characteristics of the reference individual. The TVCS, by not transforming out this effect, may more accurately reflect the occupational returns due to the average individual. On the other hand, the TVFE model produces a "pure" coefficient and so gives the return to an occupation allowing for any individual characteristics. Thus the TVCS predicts overall returns in an occupation, whereas the TVFE is more appropriate for comparing occupational differences[10].

9.2.6Age

Age is commonly included as an explanatory variable in Mincerian reduced forms as a proxy for a number of "human-capital" characteristics - experience, tenure, seniority, and so on. The age earnings profile is typically concave, reflecting the benefits of human capital accumulation at an early age[11]. This has led to the common adoption of a quadratic form for age or experience, a practice criticised by Murphy and Welch (1990) for under-estimating initial earnings growth and over-estimating the relative decline in wages of older workers.

The age profiles reported in chapter eight (figure 8.12) indicate that continuous specifications

[10] Note that the TVFE will also transform out those who remain within the same occupation in all periods - this cannot be distinguished from individual heterogeneity. Thus the TVFE model places much greater weight on changes in occupation than does the TVCS.

[11] See Berndt (1991, pp152-158). This profile is of course consistent with a number of theories: for example, job-search/segmented labour markets, forced saving, class stratification (Theodossiou(1992), Neumark (1994), Bowles and Gintis (1975)).

of the age-earnings profile are likely to perform badly. In this study a categorical variable was used to avoid imposing a specific functional form on the age profile. With enough dummy variables this is more flexible than a continuous form (although it does require more degrees of freedom, which may explain the rarity of this specification).

The profiles are given in figure 9.6, with reference age 31-35. The cross-sectional results are very appealing, both theoretically and in the apparent stability of these effects over time. This latter result is echoed in the actual age-earnings profiles which show almost no shift over time (hence the aggregation over years used to produce figure 8.12). Note that the shape of the profiles support Murphy and Welch's (1990) contention that a quadratic form would produce an excessively flat profile for young workers and an overly steep one for older workers.

However, the TVFE results make little economic sense, suggesting that, for example, in 1990 a sixteen-year-old would earn twice as much as a thirty-five-year-old doing the same job. Although results for initial years are sensible, the profile appears to be rotating about the reference age over time. This result recurs in all the TVFE studies carried out so far and only those models, and is limited to the age variables; nor is it an issue with the "pooled" or "restricted" fixed-effect models which have time-invariant slope coefficients and the expected concave shape.

Bell and Ritchie (1995a) have argued that this effect is spurious, a hitherto unreported side-effect of some models with time-varying coefficients. The reason for this apparently nonsensical result is the collinearity of time dummies (and trend variables) with variables which advance or decline in constant steps; for example, age, experience, length of residency, age of youngest child, and so on. Incremental variables incorporate an implicit trend varaible which means that the effects of time and incremental variables may not be properly separated. Most importantly, Bell and Ritchie show that there is a problem of identification with categorical variables[12]. Experiments with the data seems to suggest that the

[12] If the incremental variables are cardinal, then the model is fully identified; however, uncovering the

coefficients on age are poorly identified, and thus interpretation of the coefficients in figure 9.6a is of dubious value.

It may be argued that, in the light of these findings, TVFE models should exclude age variables. However, the TVCS model would clearly be subject to missing variable bias if the age variables were excluded, and so for comparability the age variables should be included in the TVFE regression (the curious age coefficients have little effect on the other variables, including the time dummies which are large relative to most age coefficients). Moreover, there is the small possibility that these coefficients are the genuine result of a "cohort effect", although this would require a remarkably large increase in the earnings potential of young workers which continued steadily throughout the period. Most importantly, the validity of these variables as regressors is not dependent upon being able to identify the true coefficients. Bell and Ritchie (1995a) show that although the coefficients on age are not the structural parameters, the set of age variables still contributes useful information to the model. Therefore the age dummies are included in TVFE specifications.

9.2.7Union coverage (agt)

The effect of unions on wages is a large issue which is not tackled in detail here. However, figure 9.7 depicts the coefficient on a dummy variable, set to one if earnings are affected by collective agreement. This variable is thus much wider ranging than a dummy on union membership and should avoid unmeasured spill-over effects; on the other hand, only national collective agreements are considered for this question. The net effect is that the NESPD coverage variable approximates union membership (Booth (1995)) in the proportion of individuals covered[13].

---

true coefficients still requires some manipulation of the regression results.

[13] Andrews and Bell (1995) have analysed the NES's coverage dummy using the information on local agreements collected in two years, and report that the NES coverage figures agree with other survey data when similar definitions are used.

Using this dummy as a proxy for "union effect" sidesteps a number of issues: selection bias in union presence and membership, contemporaneous membership/wage decisions, and so on (see Elliott (1991); Farber(1986, section 5); Lewis(1986) for surveys of these issues). Most importantly, there is no interaction between the union dummy and the other variables; that is, the effect of union coverage is assumed to be a simple hike in the wages of those covered, with no effect on the return of other characteristics[14]. In other words, the effect of coverage can be adequately proxied by the difference in the means of covered and uncovered workers conditional on all other features of the job or individual.

Whilst the coefficient on the dummy may be a crude measure of union influence, it has the useful feature that the effect can be followed over time; thus this variable can at least indicate the direction in which any union effect may be changing. This is of some interest given the changes in union legislation and membership since the mid-1970s but, as noted in Andrews, Bell and Ritchie (1993), studies constructing a time-series for the union effect are rare on the ground. The only other studies constructing micro time-series for the UK on a consistent basis appear to be Meghir and Whitehouse (1992) and Lanot and Walker (1993b), both using repeated cross-sections on the FES. The former found a union markup of around 10% over the period 1975-1983, and then a rise to around 17% for 1984-1986; Lanot and Walker report a markup on OLS estimates of 5% in 1978, rising to 12% by 1985. Both sets of results indicate the problems of snapshot estimates of the differential.

Figure 9.7 presents the TVFE and TVCS results. Bearing in mind the above qualifications, the "union markup" varies widely over the fourteen years. The decline in union membership over the period is reflected in the fall in the mean level of coverage. The fall in the coefficient from 1982 onwards may be due to this decline but may also reflect the anti-union legislation enacted over this period. The rise in the coefficient from 1979 to 1982 may indicate that

[14] Obviously, this is not unique to union status: a case can be made for interactions between a number of the variables: for example, industry/occupation, occupation/tenure, tenure/age, and so on.

unions were effective in maintaining wage levels over a period when the economy was undergoing a major restructuring. These results may suggest that the union effect is counter-cyclical (that is, the union wage gap is largest when demand for labour is low and smallest in a tight labour market), but the legislative and membership changes in the 1980s make this assertion difficult to prove using this data.

These results exhibit a pattern similar to those of Meghir and Whitehouse (1992) up to 1982 (if not for the subsequent three years), but the results bear little relationship to the steady rise in the coefficient from 1977 to 1985 found by Lanot and Walker (1993b).

Figure 9.7 suggests a relatively small "pure" union effect, and that it is related largely to individual ability. The cross-sectional result is exceptionally small, and is occasionally insignificant. Table 9.1 presents the t-statistics for this variable for both FE and CS results, with values significant at the 5% level in bold. In four years the agt variable is insignificant, which is unusual for this dataset where the large number of observations tends to produce high t-statistics.

Table 9.1 T-statistics for agreement variable

	1977	1978	1979	1980	1981	1982	1983	1984	1985	1986	1987	1988	1989	1990	Pool.	Rest.
TVFE	**14.954**	**9.102**	**6.525**	**9.105**	**15.148**	**18.504**	**14.689**	**11.957**	**10.383**	**8.983**	**7.494**	**5.768**	**5.512**	**2.063**	**31.127**	**32.853**
TVCS	**8.415**	1.345	1.841	**2.433**	**6.701**	**9.142**	**3.139**	**3.347**	**5.059**	**3.282**	**3.112**	-0.190	1.183	**-3.644**	**7.931**	**16.767**

Together, the results in table 9.1 and figure 9.7 contrast strongly with other findings that unions have a large effect on income of between 10% and 20%[15]. One feature of this study is the relatively high number of workplace variables, particularly in the breakdowns of occupation and industry, and it may be that the larger union markups in other are due to

[15] For example, Lewis (1986) surveys and tries to assess on a consistent basis US studies, and places most of the studies (including panel studies) within this range. Stewart (1987) using WIRS finds differentials around 10%, but this is very dependent on the characteristics of the workplace, not the worker; the differential falls to zero for some firms. Barth, Naylor and Raaum (1995) find a similar result. Murphy and Sloane (1989) using SCELI reported gaps of 15%-33% depending on allowances made for selection bias. Finally, Blackaby, Murphy and Sloane (1991) using GHS record a union gap of 28% but note that "coverage" and "membership" give substantially different results.

correlation between the variables rather than a "pure" effect[16]. Alternatively, it could be argued that the relatively small effects described here are due to collinearity between union status and, for example, occupation. In the absence of interactive dummies these competing hypotheses cannot be tested.

A final reason for the small coefficients is that the coverage variable is restricted to national agreements. Blackaby, Murphy and Sloane (1989) found that the coverage measure had a notable effect on estimates. Stewart (1987) showed that both union and employer associations had a marked effect on the apparent union wage gap, with the maximum gap being 21% but for several types of bargaining arrangement no significant effect at all. Andrews and Bell (1995), using the information on local agreements collected in 1978 and 1985, found that the inclusion of these bargains raised cross-sectional estimates by around 8%. If the lack of local agreements has this effect in all years, then the coefficients in figure 9.7, while still relatively small, are more in line with other studies.

Unusually, the TVFE model produces larger coefficients than TVCS. The implication is that the union has a larger effect when we allow for differences between individuals. This result is unexpected: Lewis (1986) argues that CS studies should represent the upper bounds on the union effect as higher union wages should lead to higher quality employees - which the TVFE model should detect. Jakubson (1991) also puts the case for a positive bias, claiming that a CS wage gap of 20% may be reduced to 5% by allowing for individual heterogeneity. Finally, Booth (1995) notes that the exaggeration of classification errors under the covariance transformation means that, theoretically at least, FE estimates cannot be larger than CS ones.

However, this argument for the "upper bound" of CS estimates ignores the potential for other errors in estimation, most importantly the potential correlation between the explanatory variables and the error terms (including the individual heterogeneity). The reverse result here

[16] Although Stewart (1987), using the establishment-level data in WIRS, finds that increasing the number of industry variables seems to make relatively difference, reducing the union wage-gap by 0.5%.

seems to suggest a negative CS bias in that individuals of relatively low "ability" are attracted to union positions through a form of adverse selection: individuals of high "ability" are encouraged to strike individual deals with employers rather than joining unions[17]. This is consistent with the stylised facts that the union effect tends to be larger for manual workers and that unions tend to have an equalising effect when different skill levels are covered (see, Farber (1986) and Lewis(1986)).

There may also be an element of selection bias. Recent studies by Lanot and Walker (1993a, 1993b) have shown that the estimated markup can vary wildly if a selection mechanism is introduced; but the effect is always to increase the markup. If the selection probability is relatively stable over time, then the TVFE model will transform this out. Thus the TVFE model may, more by accident than design, be taking account of an element of selection bias and so uncovering the true differential.

9.2.8Wages Council coverage (wbc)

Figure 9.8 shows that the effect of being in a position covered by Wages Council (WC) regulations is negative. Although this seems to imply that WC coverage reduces wages, the causation probably runs the opposite way: that the very lowest paid jobs, all other things being equal, are those most likely to be covered by the WCs. The absolute coefficients from the FE model are small for most of the time, rising at the end of the period. Those from the CS model are much larger, but fall sharply in 1982 and continue to drop until 1988 whereupon differences between the models become relatively small. The increasing wage gap between those covered and the rest of the working population is consistent with the increasing inequality in the UK labour market at the end of the 1980s (Bell(1995); Bell, Rimmer and Rimmer (1994)).

---

[17] The characterisation of unmeasured heterogeneity as "ability" here is for convenience. Similar arguments hold if "ability" is replaced by "motivation", "productivity", "nice shoes", or a number of other qualities.

These results should be treated with suspicion, as this dummy only applies to about 5% of those included in the NESPD. The FE model drops individuals with only one observation and this will reduce the number still further, which may explain the minimal coefficients for this model for most of the time. It may be that the increasing computerisation of records has been accompanied by a steady rise in the number of employees earnings under the NI limit (and possibly covered by the WCs). This could explain the increasingly strong results for the FE model, were it not for the declining mean of those covered. Moreover, the time path of the CS coefficient displays little consistent trend[18]. In short, these results do not afford a clear interpretation of the effect of WCs.

9.2.9 Time in the private sector (sector)

The proportion of individuals in the private sector increased steadily throughout the 1980s, as can be seen in figure 9.9; the relative returns to working in the private sector rose at a corresponding rate. This is consistent with the view that public sector wages tend to be counter-cyclical; that is, the private sector improves its relative pay during prosperous times (Ehrenberg and Schwarz(1986); Holmlund and Ohlsson(1992)). The TVFE model produces smaller absolute coefficients.

Interestingly, the results indicate that it is only in recent years that the private sector has become relatively better paid than the public sector - a reversal of the standard argument that public sector employment compensates for lower wages with more job security (see Ehrenberg and Schwarz(1986)). However, Rees and Shah (1992) and Bell and Ritchie (1993a) have also found a public sector premium in the hourly wage rate, for males and females respectively. Rees and Shah argue that the public sector employees work significantly fewer hours (around 7%), which could produce a private sector premium if the difference in hours is not

[18] Sudden leaps in 1982 occur in several NES statistics for no readily apparent reason.

recognised; for example, Andrews, Bell and Ritchie (1993) using a CS method on similar data to this study but with weekly wages find a public sector discount for the government sectors.

This does not explain why public sector positions should earn a higher hourly rate. Bell and Ritchie (1993a) argued that this public sector premium occurs largely in the government sector; there may be a small discount in public corporations. Given the lack of comparable "governmental" jobs in the private sector and the DE's rather idiosyncratic classification system (for example, *all* teachers are classified as "public sector"), this may be evidence of a misspecified occupation/industrial characteristic rather than a pure "sector" effect. There may also exist compensating differentials other than the job security issue (Ehrenberg and Schwarz (1986, pp1246-1251)) which can have a negative effect; for example, the disamenity experienced by dustmen or street cleaners. Finally, some authors (Hartog and Oosterbeek (1993); Rees and Shah (1992)) have argued for "comparative advantage" in the choice of sector: that people are inherently public or private sector workers by nature or early training[19]. Unfortunately, the reduced form model as specified in (9.1) is consistent with all these hypotheses and so sheds little light on the causes of the public sector premium.

9.2.10Length of time in the job (j12m)

The NES does not collect a tenure measure on a yearly basis, but it does record as a binary variable whether an individual has been in a job for at least twelve months. Figure 9.10 shows the effect of holding a job for less than one year, and it is clear that there is a discount on earnings. This is in line with human capital theory, the argument being that tenure and, by implication, experience increases (possibly firm-specific) human capital and is thus rewarded in higher wages (Mincer (1974); Coleman(1994)). However, as for the sector variables, this is not the only interpretation that can be put on these results; for example, this result would be

[19] This does raise the possibility of another source of selection bias.

expected from job-search models, increased tenure being associated with improved job matching and consequent increments to earnings (Theodossiou (1992)).

Although this variable is always significant, the effect is relatively modest, with the effective discount on earnings peaking at 3% in 1987 (TVFE specification) and generally around 2%. However, this dummy is unable to distinguish between those who have moved into jobs from unemployment and those who move from one job to another (possibly in the same company). Empirical evidence shows that those who move between jobs tend to take up a position with a higher wage, which may imply that those who stay in one post may earn less than those who move up the pay scale by changing jobs regularly. The small size of the coefficient may therefore reflect the conflation of these two opposing effects.

Note that the TVCS estimates are almost twice the size of the TVFE coefficients. This supports the view that individual characteristics influence whether people change jobs regularly, although these results cannot definitely ascribe this to heterogeneity (Cripps and Tarling (1974)) or some form of state-dependence (Phelps (1972)).

9.2.11Attrition variables

Assuming the AVs are adequate proxies for selection, then ideally the coefficients on these variables should be small for the TVFE specification at least. If both the TVFE and TVCS coefficients are insignificant, then selection bias is unlikely to be an issue; if the FE estimates are much smaller then heterogeneity plays a large part in selection, which simplifies the correction process. However, Figure 9.11 suggests that, while unmeasured characteristics may have a part to play, there remains a significant element of serial correlation and/or state dependence in selection. The coefficient on YrsIn is always very significant; the coefficients on the other two usually are. It should be noted that these attrition variables are affected by the collinearity problem discussed in 9.2.6, and also are highly correlated when the individual is in the dataset for long periods of time; they should be thus treated with some care.

The coefficients on InLast and CurrStay are almost negligible and (given that these variables are only proxies) can probably be ignored. However, although the coefficient on YrsIn appears small, the mean value of this variable is large relative to the others in the dataset and thus the total effect is relatively large[20]. The TVFE estimator appears to place a higher value on YrsIn compared to the TVCS model, suggesting that, after allowing for individual differences, complete observation history to date is even more important than recent experience.

9.3Structural shifts in the UK labour market

In this section two restrictions are placed upon the time-varying characteristic of the TVFE and TVCS models. The first is that all coefficients are constant over time (the pooled model of (9.6b) and (9.6e)); the second is that only the intercepts vary over time (the restricted model of (9.6c) and (9.6f)). Although the qualitative results on the coefficients are the same for all six models (apart from the TVFE age coefficients), regressions over the whole period unambiguously reject the hypothesis of constant slopes and/or intercepts (see table A9.1 in the chapter Appendix).

One question of interest is whether choosing a different time frame might uncover structural stability in the labour market. Identifying parameter stability with structural stability, this amounts to finding periods when the pooled or restricted hypotheses cannot be rejected.

A problem with testing this hypothesis using the TVFE model is that it needs separate means matrices to be generated for all the desired combinations of "stable" periods. However, combinations of years are easily tested in the CS framework. Therefore, despite the

[20] For example, an individual with 14 years of observations in 1990 can expect a premium of around 15% over someone making their first appearance, ceteris paribus. This would amount to a substantial premium for, for example, a woman re-entering the labour force after raising a family compared to a woman remaining in employment throughout the period.

drawbacks described in section 9.1, the TVCS model is used as a rough guide to the potential for structural shifts in the UK labour market.

Given 14 periods of observation, structural changes in the labour market can be represented in around 2^{13} ways in a simple same-not same framework. A consideration of the UK labour market simplifies this with some sensible assumptions. A wide body of evidence indicates a shift in the UK labour market at the beginning of the 1980s (see Robinson (1994), for example). The path of unemployment, peaking in 1986, also suggests a change in direction for employment prospects. Since the end of the data period (1990) does not fully reflect the economy's shift into recession (particularly with regard to expectations) a reasonable suggestion would be to look for a three-stage pattern in the UK labour market, with one shift in the early 1980s (into a period of rising unemployment and a sharp decline in manufacturing) and one in the late 1980s (with unemployment generally falling and a boom in services).

The method employed is the Chow test; that is, to run several TVCS estimates and construct F-statistics to indicate whether restrictions on the model seem justified. Allowing for structural breaks in the years {1980, 1981, 1982} and {1985, 1986, 1987} gives nine date combinations and fifteen periods to test. Table 9.2 gives the resulting F-statistics for both the pooled (all coefficients constant over the period) and restricted (intercepts allowed to vary) models, represented by "U v P" and "U v R" respectively. Degrees of freedom have not been given as in all cases these are extremely large (>200 for the numerator, >300000 for the denominator). So to test for stability over 1977-1980, for example, the pooled f-statistic is 15.45; for 1981-1986 it is 14.82.

Table 9.2 F-statistics from specification tests (cross-section)

	U v P	U v P	U v P	U v P	U v R	U v R	U v R	U v R
	80	81	82	90	80	81	82	90
77	15.45	16.85	18.12	44.08	13.02	14.85	16.45	26.65
85	19.87	11.56	6.93	17.48	12.39	9.74	5.33	8.71
86	16.80	14.82	10.75	10.18	12.96	10.43	6.48	6.90
87	20.38	18.47	14.56	6.93	13.76	11.24	7.54	5.48

The results in Table 9.2 clearly reject all the hypotheses. As the TVFE models generally give larger F-statistics, it is a fairly safe assumption that the TVFE model would also reject tests of parameter stability for all the above combinations of periods.

The rejection of the pooling hypothesis is predictable given that the model is adjusted for RPI rather than wage inflation; some trend growth in wages remains. The F-statistics for the restricted model are smaller, arising from the more general specification of this model; however, this hypothesis is also rejected by the data.

The F-statistics fall with the length of the period under review, as would be expected. However, even taking two years at a time the hypothesis of parameter stability is rejected (although the F-statistics can fall as low as 2-3). While the tests are only strictly applicable to the particular equation under review (and dependent upon the assumed normality of the error term), the suggestion is that parameter stability is a non-starter - even over short periods.

An important implication of this parameter instability is that CS studies may produce "better" models than simplistic panel studies - especially those employing differencing transformations. Compare the TVCS model of section 5.1 with the differencing approach of section 5.4. For the latter, it was remarked that the estimator forces slope parameters to be the same over the two periods being differenced. Thus the underlying assumptions of the two models are

$$\begin{array}{ll} cross\text{-}section & y_{it} = \lambda_t + x_{it}\beta_t + u_{it} \\ differencing & y_{it} = \lambda + x_{it}\beta + \alpha_i + u_{it} \end{array} \qquad \textbf{(9.7)}$$

for any two-period estimation. The CS model restricts the error term but allows parameters to vary over time, whereas the differencing model can account for individual heterogeneity but forces coefficients to be the same over any two periods. Thus there is no guarantee that the differencing panel estimator will not produce worse results than the CS model. If the restriction on parameters is carried over to more than two periods, then the validity of the differencing model is likely to decline sharply[21].

This should not be seen as a criticism of differencing models; the key point is that the parameter variability may be more important then individual heterogeneity. This issue was raised in section 2.1, when it was noted that although a panel model cannot be less efficient then a comparably specified CS model, a poorly specified panel model may perform worse than a cross-section which has different assumptions - as is the case in (9.7). This issue is peculiar to the equation being estimated, and so general conclusions about the virtues of parameter constancy versus heterogeneity cannot be drawn. However, the results presented above suggest that structural stability is not something to be assumed without some testing, and that simple CS models may be a better choice than panel specifications with ad hoc restrictions.

9.4Summary

In this chapter the first panel estimates on the NESPD have been presented and compared with cross-sectional estimates. These comparisons have supported the view that individual heterogeneity is correlated with the occupation, industry, sector et cetera of an employee, although as the TVFE model is both more efficient and has a smoothing effect on the estimates

[21] This does not invalidate the differencing approach in general; for example, the differencing model in section 5.3 (allowing for fully-varying coefficients) will out-perform the CS model.

the difference between the two models cannot be ascribed wholly to heterogeneity.

The results generally make economic sense. The wage premium on working in the public sector is unusual but is consistent with other studies using basic hourly wages. Two results are particularly noteworthy. First, the rise in private sector wages throughout the 1980s relative to the public sector seems to support a counter-cyclical hypothesis. Secondly, the decline in the union premium over the same period is consistent with the view that anti-union legislation reduced the bargaining power of unions, although this effect could also be ascribed to changes in the macroeconomy.

Finally, a crude attempt to find periods of parametric stability rejected this hypothesis in all the test cases. This has important implications for the bulk of panel models on the labour market which habitually assume structural stability over time. It may be that cross-sectional models will perform better than poorly-specified panel models which impose constant slope coefficients on the model.

Appendix A9 Regression details and summary regression results

Summary statistics for the unrestricted, pooled and restricted regressions are given in Table A9.1a (FE) and Table A9.1b (CS). Separate results for the fourteen unrestricted cross-sections are not given here, as this involves fourteen sets of summary statistics. Instead, the results for a sample year (1984, chosen for being halfway through the period) are given.

Table A9.1a Fixed-effects summary statistics 1977-1990

Fixed-effects	Unrestricted TVFE (9.6a)	Pooled (9.6b)	Restricted (9.6c)
F-test for general significance (degrees of freedom)	1011.9922 (797,907324)	11587.3083 (56, 908065)	9520.7224 (69, 908052)
R^2 Adjusted R^2 (adjustment factor)	0.4706 0.4701 (907324, 908121)	0.4168 0.4167 (908065, 908126)	0.4198 0.4197 (908052, 908121)
TSS ESS RSS Estimated variance σ^2	53436.987 25147.602 28289.385 0.031	53436.987 22270.845 31166.142 0.034	53436.987 22431.156 31005.831
Observations Restrictions Variables	1103018 194897 797	1103018 194897 56	1103018 194897 69
F-tests for models			
vs Pooled (degrees of freedom)	124.5156 (741, 907324)	-	-
vs Restricted (degrees of freedom)	119.8410 (727, 907324)	335.3523 (14, 908051)	-

Table A9.1b Cross-section summary statistics (part: 1984)

Cross-section	Unrestricted TVCS (9.6d)	Pooled (9.6e)	Restricted (9.6f)
F-test for general significance (degrees of freedom)	1573.4970 (56, 75845)	23631.9566 (56, 1102961)	22166.0247 (56, 1102948)
R^2 Adjusted R^2 (adjustment factor)	0.5374 0.5371 (75845, 75901)	0.5454 0.5454 (1102961, 1103017)	0.5925 0.5295 (1102948, 1103004)
TSS ESS RSS Estimated variance σ^2	15742.463 8460.315 7282.149 0.096	232080.576 126582.291 105498.285 0.096	221647.203 117364..033 104283.170 0.095
Observations Restrictions Variables	75902 1 56	1103018 1 56	1103018 14 56
F-tests for models			

vs Pooled (degrees of freedom)	44.0800 (728, 1102220)	-	-
vs Restricted (degrees of freedom)	26.6462 (714, 1102220)	917.9622 (14, 1102934)	-

Table A9.2 gives the results of the FE regression pertaining to the sample year 1984. All coefficients are relative to the representative categorical variables. The constant term is relative to the intercept in 1977.

Table A9.2 Time-varying fixed-effect regression results (part:1984)

Variable		Mean	Coefficient	Error	T-value	T-prob
Constant		1.000	0.0074	0.014	0.524	0.600
InLast		0.832	0.0106	0.003	4.260	0.000
YrsIn		7.202	0.0282	0.001	40.991	0.000
CurrStay		5.302	-0.0014	0.000	-3.760	0.000
reg	45	0.169	-0.0684	0.003	-24.779	0.000
reg	48	0.034	-0.0873	0.005	-17.008	0.000
reg	55	0.075	-0.1115	0.004	-28.495	0.000
reg	60	0.095	-0.1164	0.004	-30.745	0.000
reg	66	0.068	-0.1097	0.004	-26.983	0.000
reg	70	0.089	-0.1153	0.004	-29.639	0.000
reg	74	0.113	-0.1041	0.004	-28.878	0.000
reg	79	0.056	-0.1254	0.005	-26.830	0.000
reg	88	0.044	-0.1239	0.005	-24.385	0.000
reg	98	0.095	-0.0943	0.004	-22.384	0.000
agt	998	0.469	0.0229	0.002	11.957	0.000
wbc	248	0.045	-0.0173	0.004	-4.006	0.000
j12	2	0.128	-0.0282	0.002	-12.053	0.000
sec	0	0.658	-0.0175	0.003	-6.234	0.000
div	1	0.052	0.1664	0.010	17.032	0.000
div	2	0.062	0.0972	0.010	10.255	0.000
div	3	0.189	0.0561	0.009	6.096	0.000
div	4	0.113	0.0483	0.009	5.197	0.000
div	5	0.079	0.0362	0.010	3.810	0.000
div	6	0.119	0.0004	0.009	0.045	0.964
div	7	0.105	0.0668	0.010	7.050	0.000
div	8	0.080	0.0241	0.009	2.567	0.010
div	9	0.186	0.0223	0.009	2.415	0.016
age	16	0.005	-0.2778	0.015	-18.646	0.000

age	18	0.028	-0.1336	0.009	-14.890	0.000
age	20	0.042	0.0158	0.007	2.210	0.027
age	22	0.047	0.0087	0.006	1.494	0.135
age	24	0.050	-0.0079	0.005	-1.529	0.126
age	26	0.049	-0.0150	0.005	-3.262	0.001
age	30	0.096	-0.0070	0.003	-2.057	0.040
age	40	0.128	-0.0012	0.003	-0.387	0.699
age	45	0.103	-0.0074	0.004	-1.994	0.046
age	50	0.102	-0.0078	0.004	-1.859	0.063
age	55	0.100	-0.0095	0.005	-2.039	0.042
age	60	0.086	-0.0112	0.005	-2.182	0.029
age	120	0.042	-0.0056	0.006	-0.921	0.357
kos	100	0.012	0.1793	0.007	24.550	0.000
kos	122	0.075	0.1411	0.004	38.835	0.000
kos	147	0.044	0.0963	0.005	19.298	0.000
kos	156	0.009	0.1294	0.008	15.440	0.000
kos	189	0.088	0.1026	0.004	28.337	0.000
kos	211	0.065	0.1307	0.004	34.138	0.000
kos	246	0.040	0.0397	0.005	8.549	0.000
kos	254	0.030	0.0940	0.005	17.535	0.000
kos	281	0.037	-0.0968	0.005	-20.232	0.000
kos	295	0.022	-0.0625	0.008	-8.197	0.000
kos	327	0.034	0.0345	0.005	6.891	0.000
kos	385	0.052	0.0344	0.004	7.743	0.000
kos	462	0.185	0.0416	0.003	12.768	0.000
kos	477	0.041	0.0285	0.005	6.233	0.000
kos	503	0.044	0.0216	0.005	4.511	0.000
kos	533	0.108	-0.0190	0.004	-5.451	0.000
kos	540	0.019	-0.0235	0.006	-3.904	0.000

Table A9.3 lists the categorical variables used. Reference categories are marked by an asterisk.

Table A9.3 Dummy variable categories and description

Variable	Category	Description

Region		
reg	33 *	Greater London
reg	45	South East
reg	48	East Anglia
reg	55	South-west
reg	60	West Midlands
reg	66	East Midlands
reg	70	Yorkshire and Humberside
reg	74	North-west
reg	79	North
reg	88	Wales
reg	98	Scotland
Agreement		
agt	998	Earnings affected by collective agreement
agt	999 *	Earnings not affected
Wages Council board		
wbc	248	Job is covered by Wages Council regulations
wbc	249 *	Job not covered
Length of job		
j12	1 *	Current job held for over one year
j12	2	Current job held for less than one year
Sector		
sec	0	Job is in private sector
sec	3 *	Job is in public sector
Industry		
div	0 *	Farming and fishing
div	1	Energy and water supply
div	2	Other mineral and ore extraction
div	3	Metal goods, engineering and vehicles
div	4	Other manufacturing
div	5	Construction
div	6	Distribution and hotels
div	7	Transport and communication

div	8	Banking, finance and insurance
div	9	Other services
Age		
age	16	<=16 years old
age	18	17-18
age	20	19-20
age	22	21-22
age	24	23-24
age	26	25-26
age	30	27-30
age	35 *	31-35
age	40	36-40
age	45	41-45
age	50	46-50
age	55	51-55
age	60	56-60
age	120	>60 years old
Occupation		
kos	100	General management (including directorial)
kos	122	Professional and related supporting management and admin.
kos	147	Professional and related in education, welfare and health
kos	156	Literary, artistic, and sports
kos	189	Professional and related in science and engineering
kos	211	Managerial (excluding general management)
kos	238 *	Clerical and related
kos	246	Selling
kos	254	Security and protection
kos	281	Catering, cleaning, and hairdressing
kos	295	Farming and fishing
kos	327	Materials processing (excluding metal)
kos	385	Making and repairing (excluding metal)
kos	462	Process, making and repairing (metal and electrical)
kos	477	Painting, assembling, inspecting, and packaging
kos	503	Construction and mining

kos	533	Transport operating, materials moving and storage
kos	540	Miscellaneous