Andrew Barnard
Office for National Statistics

Modelling the gender pay gap in the UK: 1998 to 2006

SUMMARY

This article examines the reasons for the gender pay gap in the UK labour market by using data from the Annual Survey of Hours and Earnings (ASHE). Panel data for 1998 to 2006 is used to allow individuals to be tracked over time, using fixed-effect regression models. Regional, industrial, sectoral and other effects are investigated. The article breaks up the results using the Oaxaca method (explained in the Technical Note) to identify gender pay differentials and the trend over the time period.

The difference in the pay of men and women, commonly referred to as the gender pay gap, is a subject that attracts much interest from policy makers, researchers and the public. Studies of the gender pay gap typically aim to answer one or both of the following questions: the size of the gap and the reasons for the gap. This article will examine the latter question.

The gender pay gap is a measure of the difference between the earnings of men and women. The Office for National Statistics (ONS) regularly publishes the size of the gender pay gap in 'Patterns of Pay', and further analysis was recently undertaken by Leaker (2008). The principal source for ONS earnings estimates is the Annual Survey of Hours and Earnings (ASHE). In contrast, most empirical studies have used other sources, including the British Household Panel Survey.

This article adds to existing literature by using ASHE data for the years 1998 to 2006 using a standard wage-modelling technique. The benefit of using a wage model, compared with simple wage comparisons, is that it allows all factors to be considered simultaneously. For example, while simple statistical analysis can provide an estimate of regional differences in gender pay, the estimates will not take into account differences that may exist in the industries the individuals work in (or any other factor that affects earnings).

The article uses fixed-effect regression models, which is a technique that takes account of unobserved differences that are constant over time. For example, while

ASHE data do not include information about educational achievement (which is an important wage-related factor), education received before entry to the labour market is constant over time and is therefore controlled for using this method. The article uses the regression output to feed into an Oaxaca wage decomposition. The Oaxaca method estimates how much of the gender pay gap can be explained by differences in the observed characteristics of men and women and how much cannot. This is explained further in the Technical Note.

Why are women paid less than men?

The reason for differences in the pay of subgroups of the population is because of a combination of discriminatory and economic reasons. Determining a person's wage at the microeconomic level involves a complex interaction of several individual specific characteristics and compensating differentials, specific to individuals, jobs and workplaces. Individual characteristics are related to their productivity in the workplace and can include educational achievement and work experience. Job security and the risk of injury at work are examples of difficult to measure compensating differentials.

Discrimination occurs when one person's wage is different from another otherwise identical person's for reasons of non-productivity related characteristics, such as gender. There are several theories why discrimination exists in the labour market, including:

Box 1

The Annual Survey of Hours and Earnings (ASHE)

The Annual Survey of Hours and Earnings (ASHE) samples 0.8 per cent (1 per cent before 2007) of all employee jobs, taken from HM Revenue & Customs' pay-as-you-earn records. The Department of Enterprise, Trade and Investment conducts a similar but separate survey for employees in Northern Ireland, which enables UK-wide estimates to be made. Employers provide information about their employees, including earnings and hours worked. In 2006, ASHE had around 100,000 observations of full-time individuals who had at least one other observation in a previous time period, comprising around 62,000 males and 38,000 females.

ASHE replaced the New Earnings Survey (NES) in 2004. ASHE improved NES in the following areas:

- better coverage of employees who are lower earners
- imputation for item non-response
- weighting of earnings estimates to allow for unit non-response (not applicable to the panel analysis)

Reworking of the NES data between 1997 and 2003 using the ASHE imputation method allows for a further time series, although clearly the data do not take account of the better coverage introduced from 2004. This results in a discontinuity in the series between 2003 and 2004.

- Becker's theory of discrimination, that employers are willing to give up some profits to pay for their taste for discrimination
- judging individuals based on the average characteristics of a group they belong to, known as statistical discrimination. An example of this is the assumption that a woman aged 30 will go on maternity leave in the near future. The cause of this is imperfect information. This leads employers to make statistical assumptions based on averages – whether accurate or not – to reduce their (perceived) risk
- crowding models which lead to occupational segregation. For example, some occupations may be more attractive to women and, therefore, all other things being equal, the increased supply of labour will lead to reduced wages within those occupations

Some discrimination occurs to people before they enter the labour market, whereas other discrimination occurs within the labour market. For example, the crowding model could be related to unequal promotion of some occupations among the sexes. Expectations could also be a factor since lower expectations of wages can influence career choices.

Using panel data to model the gender pay gap

Analysis of the gender pay gap is often presented in terms of average differences in the pay of men and women with comparison to one other variable. For example, analysis will often quote the average difference in men's and women's wages in an industrial sector. However, this has limitations, since the factors that influence a person's wage do not act in isolation. Regression techniques enable wages to be modelled using many variables at once, which can give better estimates of the effect of each factor on earnings.

Wage modelling falls into two broad categories:

- cross-sectional analysis, using data for one time period, and
- panel (or longitudinal) analysis, where the same individuals are analysed over many time periods

The are many advantages to using panel data rather than cross-sectional, including increased degrees of freedom, reduced problems of data multicollinearity and controls for time-invariant variables which cannot otherwise be included (Hsiao 2003). The last point is useful when dealing with labour market data. For example, education received before entry to the labour market would be controlled for in a fixed-effect wage equation model even if the data set did not include individual educational attainment information. Application of the same principal to other individual characteristics that do not change over time is possible.

ASHE variables used for wage modelling

Since most ASHE data come from the company payroll, it is an excellent source for earnings estimation – many other sources rely on respondents' answers to earnings questions which can be inaccurate. A downside of ASHE data is that they lack information about individual characteristics which would be present in a perfect wage model. For example, ASHE does not hold information on motivation and yet this will have an effect on a person's wage. Also, ASHE does not record whether an individual has returned to work from unemployment or inactivity (for example, being on maternity leave) but these factors are known to matter. Nevertheless, ASHE does have several variables which can be used directly or as proxies for individual characteristics and compensating differentials. These include age (an imperfect but reasonable proxy for work experience), tenure, occupation, industry of employment, region and coverage of the wage by a collective bargaining agreement.

The analysis considers only individuals in full-time employment. ASHE defines part-time as less than 30 basic weekly paid hours (except for teaching professionals which is less than 25 hours). Basic weekly paid hours refers to the weekly average number of hours paid at the basic rate of pay during the pay period that includes the survey reference date. The analysis excludes people who work part-time because of differences in their characteristics compared with people who work full-time. However, since excluding part-time individuals could lead to selection bias, the analysis was conducted with part-time individuals included, for comparison, with no visible differences to the results.

The data set employed by this study uses ASHE and reworked New Earnings Survey (NES) data for 1998 to 2006 (see **Box 1**). Merging the data for each year creates a panel, using the unique identifiers to distinguish individuals. Individuals identified as potentially sharing a National Insurance number, where individuals have a temporary number and individuals with second jobs are removed so the analysis covers main jobs only.

Model and methodology used

The results presented are based on a log-wage equation for female and male hourly earnings, based on a standard method, using a panel model approach:

$$\ln(w_{it}) = \beta X_{it} + \lambda_t + \alpha_i + \mu_{it} \quad (1)$$

where w_i is hourly earnings (for individual i), X_{it} is a vector of conditioning variables, β represents the rates of return to characteristics where $\beta_t = \beta$ for all t, λ_t and α_i are the coefficients on time and individual specific dummies, and μ_i is the disturbance term. The wage is logged as this has the useful property of causing the resulting coefficients to be the per cent effect on earnings.

It is also possible to relax the restriction to the rate of return to characteristics to allow it to change over time, so:

$$\ln(w_{it}) = \beta_t X_{it} + \lambda_t + \alpha_i + \mu_{it} \quad (2)$$

The benefit of equation (2) (allowing β to vary over time) is that it allows individuals who do not change status over time to contribute to the regression, whereas in (1) they do not (Bell and Ritchie 1998). A potential problem with equation (2) in practice is that it needs estimations of a far higher number of coefficients than equation (1), by a factor of the number of years in the panel data set. This can increase the size of the data matrix. Therefore, equation (1) was regressed over all t and for subsets of t to allow tracking of the estimated coefficients over time.

Estimation of the model takes place for both male and female workers to find out whether the coefficients of the variables differ, showing a difference in the rate of return of working in, for example, a particular occupation or sector. Any difference on the rates of return would suggest gender differences in the workings of the UK labour market.

The gender pay gap can be split into two parts. The first is the 'explained' part, because of differences in the characteristics of men and women. The second is the 'unexplained' part, which refers to differences in the rate of return to characteristics. Using the model can assign how much of the gender pay gap is caused by each part. For further explanation, see the Technical Note on the Oaxaca Decomposition. Gender-specific differences in the rate of return to characteristics suggest that wage decisions by employers were made for non-productivity reasons. Therefore, using this method, the unexplained part gives a proxy for the level of discrimination which can be assessed over time. However, the unexplained part will inevitably include the effect of compensating differentials and individual characteristics that are not included within the model's specification. In this case, since ASHE has a shortage of individual characteristic variables, the model will lead to a large unexplained part not because of discriminatory reasons. Also, there are strong arguments that the 'explained' part could carry discrimination, for example, pre-labour market discrimination leading to women concentrating in lower-paid occupations.

Results

In the earnings model used, the dependent variable is the natural logarithm of hourly earnings excluding overtime, adjusted to the retail prices index to the base year 1987. The model includes all males and females employed full-time, with at least two observations for the years 1998 to 2006.

Removal takes place for individuals whose earnings are affected by absence.

The explanatory variables used are:

■ occupational dummies based on the Standard Occupational Classification 2000 (SOC2000) one-digit groups. As the data were recorded using the previous SOC90 before 2002, mapping of observations to SOC2000 before 2002 allows comparison over time. See the Technical Note for further details

■ industrial dummies, based on the Standard Industrial Classification 2003 (SIC2003). Because there are 17 sub-groups in SIC2003, these are arranged into nine groups. See the Technical Note for further details

■ regional dummies, based on Government Office Region of workplace

■ sector dummy, whether the employer is in the public or private sector

■ coverage by collective bargaining agreement. It is worth noting that the collective bargaining agreement variables created using ASHE have discontinuities caused by changes to questionnaires and other reasons

■ tenure dummy, whether the time spent in the current job is greater or less than one year

■ age. Here the model uses two approaches, with the first a quadratic function of the age variable and second, dummy variables using age intervals

■ selection dummy variables are included in the model because of the usual selection problem

Table 1
Full-time workers with at least two observations and no periods of pay affected by absence, excluding overtime

United Kingdom

	Observed		Mean weekly wage (£) (2006 prices)		Median weekly wage (£) (2006 prices)		Mean hourly wage (£) (2006 prices)		Median hourly wage (£) (2006 prices)		Median hourly wage ratio (%)	Mean hourly wage ratio (%)
	Male	Female	Male	Female	Male	Female	Male	Female	Male	Female	Female/ Male	Female/ Male
1998	69,864	39,114	484.52	370.59	397.31	321.92	12.75	10.26	10.28	8.70	80.5%	84.6%
1999	69,529	39,364	496.58	384.93	406.36	332.88	13.08	10.67	10.51	9.00	81.6%	85.7%
2000	67,526	38,635	503.56	391.18	408.32	335.18	13.25	10.83	10.57	9.07	81.7%	85.8%
2001	67,019	39,081	524.23	410.89	422.63	351.05	13.81	11.36	10.92	9.43	82.3%	86.3%
2002	67,109	39,589	544.97	423.68	432.71	360.62	14.37	11.69	11.24	9.73	81.3%	86.5%
2003	67,487	40,548	545.92	427.88	436.30	364.99	14.36	11.81	11.24	9.89	82.2%	88.0%
2004	65,874	39,828	537.01	430.98	435.03	366.59	14.11	11.87	11.22	9.91	84.1%	88.3%
2005	66,189	41,288	547.28	439.47	440.74	373.43	14.44	12.14	11.41	10.14	84.1%	88.9%
2006	62,181	38,156	560.53	449.72	451.29	383.30	14.73	12.39	11.63	10.41	84.1%	89.5%

Note:
Total observations: 958,381 (602,778 male; 355,603 female).

Source: Annual Survey of Hours and Earnings

Inclusion of industrial and occupation dummies identifies some of the compensating differentials in the model. There are arguments against including occupational variables in wage modelling, since they may be endogenous to people's wages. This is because a person's choice of occupation may be influenced by average wages on offer by occupations.

There are three types of regressions run on the data:

- cross-sectional regressions for each year 1998 to 2006 (CS)
- a panel fixed-effect regression for the years 1998 to 2006 (PFE)
- three separate fixed-effect panel regressions for the years 1998 to 2000, 2001 to 2003 and 2004 to 2006 (PTV)

Separate regressions take place for men and women.

Nearly all observations exceed the 5 per cent significance level because of the large number of observations; this includes t-statistics and F tests. A standard (Hausman) test supported the use of the fixed effects model.

Females account for just over one-third of the total observations between 1998 and 2006 (see **Table 1**). Over this period, the female to male median hourly wage ratio, excluding overtime, has risen from 81 to 84 per cent, meaning the gender hourly wage gap has fallen from around 19 to 16 per

cent. Leaker (2008) describes alternative measures of gender pay differentials for the UK.

The effect of age on earnings is as expected, that is, as age increases there is a positive but decreasing effect on earnings. Experiments with a quadratic age function and age-banded dummies provided similar regression estimates and this article presents the latter.

Comparing the age effect for females and males shows that men progress faster than women until about age 21. Earnings then increase at a similar rate until age 40, after which the increase in females' earnings is steeper than males' (see **Figure 1**).

Bell and Ritchie (1998), using NES data for the period 1977 to 1994, found that females had a flatter earnings profile than males. They argue this could be because age represents a better proxy for work experience for men, since women are more likely to take career breaks. It is therefore useful to compare the results of PFE regression with CS regression to assess for any bias. For example, many high-earning women may return to the labour market later in life. All cross-sectional results provide flatter earnings profiles for women than men, supporting the argument that women returning to the panel may be causing an effect to the analysis, at ages 45 and over.

The crude tenure variable, which takes account of people who have been in their

job for less than one year, highlights little difference between males and females. The analysis highlights that there is a negative effect of approximately 3 per cent of earnings for those people who have been in their job for less than one year. This is a small increase on previous studies, although PTV regression does not pick up any trend in the effect over the time period.
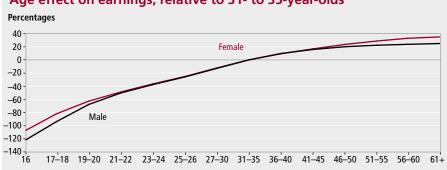
The results show that, on average, people working in the public sector earn more than otherwise identical people working in the private sector and the effect is larger for women than for men. This is often referred to as the 'public sector premium'. Differences in understanding and applying pay equality legislation are attributed to causing part of the public sector premium. For example, working conditions offered by public sector employers are often more flexible than those in the private sector, and this could benefit women more than men. Preliminary work by Chatterji *et al* (2007) supported this, implying that family friendly policies in the public sector translated to higher public sector wages, especially for women. Comparing the results of the PTV regressions, it appears that the public sector premium for both men and women grew in 2004–06.

A feature of all the results is that the regional earnings effect of not working in London is much larger for women than men (see **Figure 2**). The regional effect on female earnings is always more negative than on males, relative to London (other than in the South East in 2004–06, PTV results). This supports theories of reduced labour mobility within the female workforce which leads to labour supply imbalances and therefore reduced earnings in some areas (namely, all bar London). Since London has a much larger and integrated job market than other regions, it is likely to lessen labour mobility issues.

The results of the PFE regression highlights that men in the South East earn 4 per cent less than their London counterparts (all others things being equal), whereas women earn 7 per cent less. A possible explanation is that while men benefit from proximity to London, the benefits do not pass as well to women. Interestingly, the PTV analysis shows that women in the South East significantly improved their position compared with women in London over the time period. There is no graphical presentation of the results from the PTV regression because of the large numbers of coefficients. These are available on request.

Regional Trends 2008 provided estimates

Figure 1
Age effect on earnings, relative to 31- to 35-year-olds

Figure 2
Regional effect on earnings, relative to London

of regional gender earnings gaps using data from the Survey of Personal Incomes (managed by HM Revenue & Customs). The largest proportional difference in gender pay for the period presented (2004/05) was in the South East and the smallest was in Wales. At first this appears to contradict the results presented in Figure 2, as this shows the two regions have similar differences in gender pay. However, this demonstrates the limitation of simple estimates of gender pay differentials compared with estimates (as in this article) which take account of some of the reasons which influence a person's wage.

A second regional effect highlighted by the PTV regression results is improving wages in Wales, compared with wages in London. The relative improvement for men over the time period exceeds that for women. This could contribute to Wales having the largest regional gender difference in earnings coefficients (see Figure 2). The regeneration of Wales, especially in Cardiff, may explain this improvement. An alternative or added explanation is that labour market interventions such as the National Minimum Wage and New Deal have had a disproportionate effect in Wales.

Males living in the South West of England have improved their position compared with males in London in recent years, although this improvement is not seen in the female group.

Focusing on occupational effects in the PFE regressions, differences are observed in the gender returns to working within Managers and senior officials, Professional occupations and Elementary occupations, compared with Administrative and secretarial occupations. In other occupations there is little difference between the female and male returns (see **Figure 3**). Males receive a 2 per cent greater premium for working in Managerial and senior positions whereas females receive a 2 per cent greater premium for working in Professional occupations. A possible explanation is that individuals working within Professional occupations find it easier, and less damaging to their future earnings (if at all), to take career breaks than those working in Senior official occupations. For example, Teaching professionals and Chartered accountants (subsets of Professional occupations) may be more flexible careers than Directors and chief executives of major organisations and Purchasing managers (subsets of Senior officials). A feature of the PTV regressions is that the occupational earnings effects are smaller in the 2001 to 2003 period for both males and females.
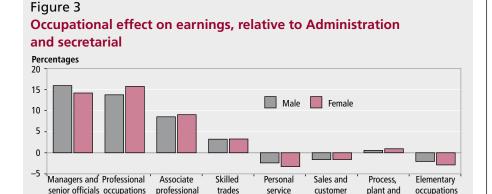
The results of the PFE regression show a larger wage premium for males in manufacturing (the industrial effect) (see **Figure 4**). This could be a result of women

working in less skilled manufacturing positions, or could be evidence of discrimination in that sector. The former argument is backed up loosely through evidence such as that presented by the House of Commons Trade and Industry Committee (TSO 2007). This reported that women make up 25 per cent of those reading manufacturing-related degrees, 3 per cent of modern apprentices in manufacturing and engineering and 6 per cent of professional engineers. There is also a larger wage premium for males in Finance, Mining and quarrying, Energy and water and Agriculture and fishing. In fact, the only industry to show a slightly higher wage premium for women was in the group Other services (public), which could be because of the public sector premium effect.

To test the effect of the selection dummies, the PFE regression was undertaken with the selection dummies removed. The effect was to make the tenure coefficient larger (1 per cent greater impact on wages) and all age coefficients were larger, especially those for the lowest and highest age bracket dummies. This is not surprising, since the likelihood of going into unemployment or economic inactivity is highest when changing jobs, or when young or old (Bell and Ritchie 1998).

Decomposing the results

Finally, an Oaxaca type decomposition is applied to the results of the CS, PFE and PTV regressions: the last named is presented in **Table 2**. See the box in the Technical Note for further information about how to interpret these results, including a detailed explanation of the weighting methods used. A simple interpretation of the decomposition is that the male and female weightings are the upper and lower bounds of the part of the wage difference accounted for by explained and unexplained reasons. The pooled weighting is therefore a midpoint (calculated according to the Oaxaca-Ransom method) between the upper and lower bounds. Figures in italics show the proportion of the total explained and unexplained.

The total weighted effect falls over time, especially from 2001–03 to 2004–06. This shows that the modelled gender wage gap falls over the time period. All weighting methods display a fall in the unexplained part over all time periods, while the explained part rises and falls. The pooled weighting method shows that almost two-thirds of the wage gap is because of reasons unexplained and one-third for reasons

Figure 3
Occupational effect on earnings, relative to Administration and secretarial

Figure 4
Industrial effect on earnings, relative to Real estate, retail and wholesale
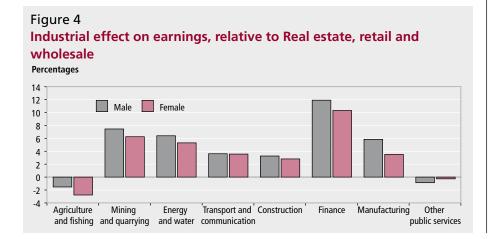
Table 2
The Oaxaca decomposition

Weighting method		1998–2000		2001–03		2004–06	
		Weighting	Proportions (%)	Weighting	Proportions (%)	Weighting	Proportions (%)
Male	Explained	0.027	17	0.021	14	0.026	20
	Unexplained	0.134	83	0.127	86	0.100	80
Female	Explained	0.021	13	0.020	13	0.015	12
	Unexplained	0.141	87	0.129	87	0.110	88
Pooled	Explained	0.053	33	0.048	33	0.045	36
	Unexplained	0.109	67	0.100	67	0.080	64
Total		0.161		0.148		0.126	

explained. The unexplained element is large, primarily because of a lack of individual characteristics variables in ASHE; however, this is constant over time. Therefore, the falls in the unexplained component could be caused by a reduction in discrimination over the time period, although other factors that cannot be measured could also be the reason.

Looking at the decompositions in detail, it is clear that most of the fall in the unexplained part is because of falls in the occupational and age terms and because of large falls in the constant term. Interestingly, the sector terms show increases over the time periods. The small falls seen in nearly all explained parts are because of falls in the age and occupation elements.

Conclusions

Reductions in the gender pay gap since 1998 can be attributed mainly to unobservable differences between men and women. Due to limitations in the data, it is not possible to say how much of the decrease is due to a reduction in discrimination and how much is because of other unobservable factors that cannot be identified in ASHE.

Of the observable factors, the age, region, occupation, industry and sector variables have a significant impact on earnings for men and women. In particular, women's wages are more dispersed by region than men's and, save for the exceptions of South East and Wales, there is little change over the period. This supports theories of reduced female labour mobility which result in lower female earnings in all regions except London. Senior official occupations, which are among the highest paid, benefit men's wages more than women's, although Professional occupations benefit women more than men. This could be due to more flexibility within the Professional occupations, compared with Senior officials. It is also clear that there is a benefit for working within the public sector, and the effect is greater for women than men.

CONTACT

✉ elmr@ons.gsi.gov.uk

REFERENCES

Bell D and Ritchie F (1998) 'Female earnings and gender differentials in Great Britain 1977–1994', *Labour Economics* 5: pp 331–357.

Chatterji M, Smith P and Mumford K (2007) 'The Public-Private Sector Gender Wage Differential: Evidence from Matched Employee-Workplace Data', Institute for the Study of Labor Discussion Paper.

Hsiao C (2003) 'Analysis of Panel Data, Second Edition', Econometric Society Monographs.

Jasso G and Webster M (1999) 'Assessing the Gender Gap in Just Earnings and Its Underlying Mechanisms', Social Psychology Quarterly: vol 62, no 4.

Leaker D (2008) 'The Gender Pay Gap in the UK', *Economic & Labour Market Review* 2(4), pp 19–24.

Mincer J (1974) 'Progress in Human Capital Analysis of the Distribution of Earnings', *NBER Working Papers*: 0053 National Bureau of Economic Research.

Oaxaca R (1973) 'Male Female Wage Differentials in Urban Labor Markets', International Economic Review 14(3): pp 693–709.

Oaxaca R and Ransom M (1994) 'On Discrimination and the Decomposition of Wage Differentials', *Journal of Econometrics*, 61, pp 5–21.

Office for National Statistics 'Patterns of Pay' at www.statistics.gov.uk/statbase/product.asp?vlnk=14123

Office for National Statistics (2003) 'Census 2001 – London and Its People' at www.statistics.gov.uk/census2001/profiles/commentaries/london.asp

Office for National Statistics (2008) *Regional Trends* 40 at www.statistics.gov.uk/regionaltrends40/

Rubery J and Smith M (2006) '*The UK Gender Pay Gap: Recent Developments*', European Expert Group on Gender, Social Inclusion and Employment.

The Stationery Office (2007) '*Better Skills for Manufacturing*', House of Commons Trade and Industry Committee.

Walling A (2007) 'Understanding statistics on full-time/part-time employment', *Economic & Labour Market Review* 1(2), pp 36–43.

TECHNICAL NOTE

The Oaxaca decomposition

To find out whether differences in pay are caused by differences in characteristics or differences to rates of return to the characteristics, an Oaxaca (1973) decomposition method is used. This is shown as:

$$\overline{\ln(w_m)} - \overline{\ln(w_f)} = \sum_{i=1}^{n} (\overline{X}_{mi} - \overline{X}_{fi}) \cdot \beta_{mi} + \sum_{i=1}^{n} (\beta_{mi} - \beta_{fi}) \cdot \overline{X}_{fi}$$

Where $\overline{\ln(w_m)}$ and $\overline{\ln(w_f)}$ are the average (naturally logged) wages for males and females respectively, $\overline{X}_{mi}$ and $\overline{X}_{fi}$ are the means of the observations for individual i, and β_{mi} and β_{fi} are the estimated coefficients for individual i for males and females.

In the equation, the first term on the right-hand side represents the part of the gender wage difference that is caused by the difference in observed human capital between males and females. For example, the difference in experience (in this article proxied by age) and differences with any non-productivity related variables included in the analysis (differences in occupational, industrial, sectoral distributions). This is usually referred to as the explained, or observable, part of the decomposition.

The second term represents the part of the gender wage difference which cannot be explained by differences in human capital, or by any of the other variables used. Therefore, this is attributed to differences in the rate of return to wage determining variates. This is referred to as the unexplained, or unobservable, part.

It is useful to interpret the decomposition as follows:

	Possible non-discriminatory components	**Possible discriminatory components**
Explained	Differences in compensating differentials. Differences in individual characteristics.	'Pre' and 'within' labour market discrimination that prevents women from improving their human capital.
Unexplained	Omitted variables.	Discrimination that prevents women from obtaining the same rate of return to the wage determining variates.

The Oaxaca method presented above assumes that the 'post-model' wage for both males and females (that is, the wage rate for an identical male or female) equals the 'pre-model' wage for men (W_m). In Table 2, this is labelled as the 'Male' weighting method. It is also possible to rearrange the equation to enable other scenarios to be considered. In the decompositions presented, the other scenarios considered for the post-model wage structure are as follows:

• the same as the average female pre-model wage ('Female'), and

• weighted according to the Oaxaca-Ransom method (Oaxaca and Ransom 1994) ('Pooled')

Additional weighting methods, not presented for sake of space, include:

• an average of male and female pre-model wages ('Average')

• a weighted average of male and female pre-model wages, according to the number of males and females in the sample ('Weighted Average')

Mapping SOC90 to SOC2000

SOC2000 group	**Original SOC90 two-digit groups**
Senior officials	10, 11, 12, 13, 14, 15, 16, 17, 19
Professional occupations	20, 21, 22, 23, 24, 25, 26 27, 29
Associate professional and technical occupations	30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 60, 61, 70, 71
Administrative and secretarial occupations	40, 41, 42, 43, 45, 46, 49
Skilled trades occupations	50, 51, 52, 53, 54, 55, 56, 57, 58, 59
Personal service occupations	63, 64, 65, 66, 67
Sales and customer service occupations	72, 73, 79
Process, plant and machine operatives	80, 81, 82, 83, 84, 85, 87, 88, 89
Elementary occupations	86, 90, 91, 92, 94, 95, 99

Grouping for SIC2003

Industrial groups used in this article are combined into the following nine groups, showing SIC2003 constituent letter groups:

• Agriculture and fishing (A+B)

• Mining and quarrying (C)

• Manufacturing (D)

• Energy and water (E)

• Construction (F)

• Transport and communication (I)

• Finance (J)

• Real estate, retail and wholesale (G+H+K+O+P+Q)

• Other public services (L+M+N)