A New Decomposition of the Wage Differential between Genders

Dilaka Lathapipat[†]

November 3, 2006

Abstract

In this work we argue that the Oaxaca-Blinder decomposition technique incompletely decomposes the wage difference between two comparative groups into a portion that is explained by differences in characteristics, and a portion that is attributable to differences in the valuation of these characteristics. As a result, studies on earnings inequalities that rely on the traditional methodology tend to underestimate the explained portion of the earnings differential. To rectify this problem this paper proposes for the first time an alternative decomposition technique that is an extension of the traditional method. To illustrate the concept, the proposed methodology is applied to the study of gender wage inequality using Australian 2001 data from the HILDA survey. The traditional Oaxaca-Blinder decomposition is also applied to the same data so that the degree of underestimation can be easily inferred.

Keywords: earnings inequality, gender inequality

JEL Classification: J31

* Paper submitted in fulfilment of the requirements for the course EMET8036: Economics Seminars.
† PhD scholar in Economics, Research School of Social Science, The Australian National University.
I thank Professor Bruce Chapman for his supervision, Dr. Robert Breunig and Professor Martin Richardson for valuable comments on an earlier draft. Any errors are my own responsibility.
E-mail: dilaka.lathapipat@anu.edu.au

INTRODUCTION

The Mincer Earnings function and its associated parabolic wage equation (Mincer, 1974) has been widely used in empirical labour economics. One popular application of the equation in the last 30 years has been in studies relating to gender wage inequality. Typical approaches involve estimating separate equations for males and females and decomposing the observed differences into “explained” and “unexplained” components using a decomposition technique first introduced by Oaxaca and Blinder (1973).

This paper argues that past studies based on the traditional Oaxaca-Blinder decomposition obtained very misleading estimates of the explained and unexplained portions of the wage difference. This is primarily due to an incomplete decomposition of unobserved productive characteristics, as well as returns on post-schooling human capital investment, that are embedded within certain key regression parameters. Specifically, the unobserved characteristics are the average post-schooling investment behaviours, which are potentially very different between males and females.

In order to illustrate the conceptual framework, the Mincer Earnings function is explored in detail. A method for obtaining estimates for the parameters in the post-schooling human capital investment function is then proposed, together with a modified version of the Oaxaca-Blinder decomposition that accommodates these new estimates. For illustration, the framework is then applied to the 2001 Australian data from the HILDA Survey and the resulting decomposition into the explained and unexplained portions are presented together with those obtained using the traditional decomposition method.

LITERATURE REVIEW

In studies of gender wage differential (see for examples: Chapman and Mulvey, 1986; Reiman, 1998; and Crockett and Preston, 1999), most labour economists have adopted an approach involving the estimation of separate male and female Mincer parabolic wage equations (Mincer, 1974) controlling for a range of relevant characteristics such as occupation and industry of employment as well as other demographic and personal characteristics. The wage differential is then decomposed into "explained" and "unexplained" portions using a widely popular decomposition method introduced by Oaxaca and Blinder (1973). First, the following two equations are estimated for males (*m*) and females (*f*) respectively:

$$\ln \hat{Y}_i^m = \hat{\beta}_0^m + X_{p,i}^m \hat{\beta}_p^m + X_{c,i}^m \hat{\beta}_c^m \quad (1)$$

$$\ln \hat{Y}_i^f = \hat{\beta}_0^f + X_{p,i}^f \hat{\beta}_p^f + X_{c,i}^f \hat{\beta}_c^f \quad (2)$$

The vector $X_{p,i}$ comprises productive characteristics such as the schooling and the quadratic experience variables while the vector $X_{c,i}$ contains control variables that are thought to affect the wage rate. Afterwards, the following decomposition is used to determine the explained and unexplained components of the sample average gender wage difference:

$$\overline{\ln Y^m} - \overline{\ln Y^f} = (\overline{X^m} - \overline{X^f})\hat{\beta}^m + \overline{X^f}(\hat{\beta}^m - \hat{\beta}^f) \quad (3)$$

Note that for compactness the means of the productive and control variables are combined into a single vector $\overline{X}$, and the vector of estimated coefficients $\hat{\beta}$ are also similarly combined. The first term on the right-hand-side represents the portion of the difference in wages attributable to differences in average characteristics – hence the “explained portion”, while the last term represents the difference in returns paid to these characteristics – hence the “unexplained portion”.

For the case of Australia, using 1982 data and hourly earnings, Chapman and Mulvey (1986) found the average raw wage differential to be around 17% between males and females, with the explained portion accounting for only 21% of the difference and the unexplained portion accounting for 79%. More recent studies have found the gap to be around 12%-14%, with the explained portion accounting for only 27%-32% and the unexplained portion 68%-73% (see Reiman, 1998 and Crockett and Preston, 1999). These studies therefore conclude that around three quarters of the gender wage gap may be attributable to the undervaluation of female skills.

In this paper I intend to point out why all of these studies could be underestimating the explained portion of the wage gap by overlooking potentially important information regarding productive characteristics embedded within several regression coefficients. To see this we need to look beyond the surface of the Mincer parabolic wage equation and analyse the components of these few key coefficients more carefully.

An Analysis of the Mincer Human-Capital Earnings Function

In this section I follow Mincer closely in the derivation of his celebrated Human-Capital Earnings Function and its associated parabolic wage equation. Consider C_j the amount of net “time-equivalent” human capital investment *j* years into work experience, and E_j the corresponding gross earnings capacity of a representative individual. Let $k_j = C_j / E_j$ be the net human capital investment ratio and r be the average marginal return on the investment, the relationships can be shown as:

$$E_j = E_{j-1} + rC_{j-1} = E_{j-1}(1 + rk_{j-1}) \quad (4)$$

By recursion of equation (4) the gross earnings capacity can be written as:

$$E_j = E_0 \prod_{t=0}^{j-1} (1+rk_t) \quad (5)$$

Equation (5) can be log-linearised and using the approximation that $\ln(1+rk_t) \simeq rk_t$ for small rk_t:

$$\ln E_j = \ln E_0 + r\sum_{t=0}^{j-1} k_t \quad (6)$$

Let observable net earnings in year *j* after schooling (or equivalently in the *j*th year of work experience) be given by $Y_j = E_j - C_j = E_j(1-k_j)$. Hence:

$$\ln Y_j = \ln E_j + \ln(1-k_j) \quad (7)$$

Substitute (6) into (7) yields:

$$\ln Y_j = \ln E_0 + r\sum_{t=0}^{j-1} k_t + \ln(1-k_j) \quad (8)$$

Assuming that during the schooling period, all of the time available is devoted to education and hence $k_t = 1$ and the average marginal returns during the s years of schooling is r_s. This assumption leads to $\ln E_0 = \ln Y_0 + r_s s$ for some initial earnings capacity Y_0 before any formal human capital investments are undertaken. For notational clarification, the summation in the second term on the RHS of (8) captures the periods after formal full-time schooling investment. Therefore:

$$\ln Y_j = \ln Y_0 + r_s s + r\sum_{t=0}^{j-1} k_t + \ln(1-k_j) \quad (9)$$

For convenience treat the investment and earnings functions as continuous functions of time and, changing the subscripts for the number of years of work experience from *j* to *t*, write (9) as:

$$\ln Y_t = \ln Y_0 + r_s s + r\int_{j=0}^{t} k_j dj + \ln(1-k_t) \qquad (10)$$

Assume also that the net investment ratio takes a linearly declining profile and denote by k_0; the initial net investment ratio at the beginning of work experience, and by T; the total periods of positive net investment. Then:

$$\ln Y_t = \ln Y_0 + r_s s + r\int_{j=0}^{t} \left(k_0 - \frac{k_0}{T} j\right) dj + \ln\left(1 - k_0 + \frac{k_0}{T} t\right) \qquad (11)$$

Integrating the term in the integral and using second-order Taylor expansion for the right-most term around 1 in (11) and rearranging gives:

$$\ln Y_t = \left(\ln Y_0 - k_0 - \frac{k_0^2}{2}\right) + r_s s + \left(rk_0 + \frac{k_0}{T} + \frac{k_0^2}{T}\right) t - \left(\frac{rk_0}{2T} + \frac{k_0^2}{2T^2}\right) t^2 \qquad (12)$$

Equation (12) forms the basis of the parabolic earnings function proposed by Mincer. The first term on the RHS is interpreted as the intercept of the regression equation, while the third term and the negative of the fourth term in parentheses are the slope coefficients on the experience and experience-squared variables respectively.

When looking at (12) as a series of log-earnings for an individual, it is clear that there is a trade-off between lower initial earnings from investing in post-schooling human capital investment, and higher rate of growth of that wage rate with experience. Note also that a higher initial investment ratio also implies a more concave log-earnings profile. However, since we do not have access to sufficiently long panel data that tracks individuals' earnings throughout their lives; we have to rely on cross sectional data instead. Analyses based on cross sections, at least for the younger sample around the early to mid part of their careers, should allow valuable inferences.

The Variance of the Mincer Earnings Function

The analysis of the variance of log-earnings in this section is based on Mincer and is also shown in Heckman (2005). Using equation (8), substitute for k the linearly declining net investment function and making the simplifying approximation that $\ln(1-k_t) \simeq -k_t$ for small k_t. Let t denote years of work experience, then after some rearranging:

$$\ln Y_t = \ln E_0 + k_0 \left\{ r\sum_{j=0}^{t-1}\left(1-\frac{j}{T}\right) - \left(1-\frac{t}{T}\right)\right\}$$

Assuming that only the initial earning potential $\ln E_0$ and initial net investment ratio k_0 vary in the population, the variance of log-earnings is given by:

$$Var(\ln Y_t) = Var(\ln E_0) + \left\{ r\sum_{j=0}^{t-1}\left(1-\frac{j}{T}\right) - \left(1-\frac{t}{T}\right)\right\}^2 Var(k_0) + 2\left\{ r\sum_{j=0}^{t-1}\left(1-\frac{j}{T}\right) - \left(1-\frac{t}{T}\right)\right\} Cov(\ln E_0, k_0) \quad (13)$$

Ignoring for the moment the covariance between k_0 and $\ln E_0$, the variance of log-earnings is minimised and is equal to $Var(\ln E_0)$ when:

$$r\sum_{j=0}^{t-1}\left(1-\frac{j}{T}\right) = \left(1-\frac{t}{T}\right) \quad (14)$$

The LHS of (14) can be manipulated to yield:

$$r\left(t-\frac{t(t-1)}{2T}\right) = \left(1-\frac{t}{T}\right) \quad (15)$$

When T is large, it is easy to see that the variance minimising experience level t^* approaches $1/r$. To be more precise, (15) can be expanded to obtain:

$$\frac{r}{T}\left[\frac{t^2}{2} - \left(T+\frac{1}{2}+\frac{1}{r}\right)t + \frac{T}{r}\right] = 0 \quad (16)$$

The variance minimising solution of experience level is thus:

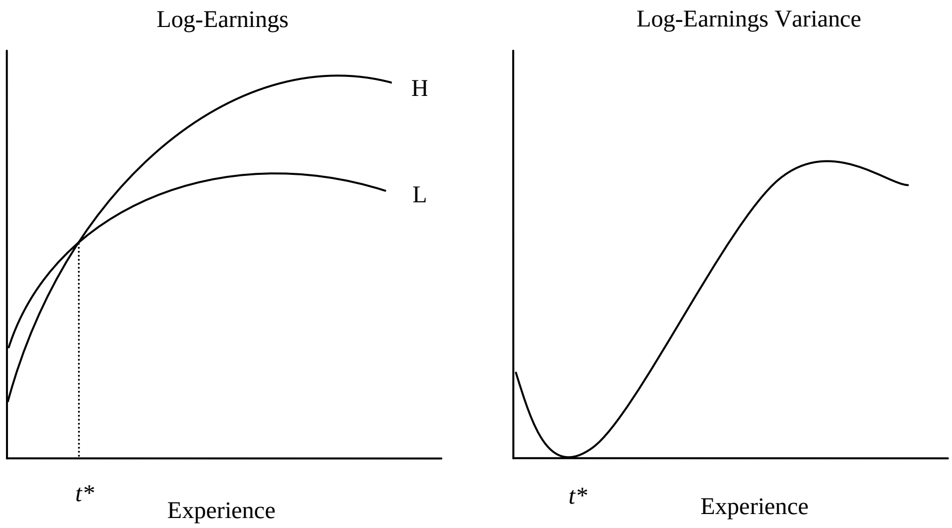
$$t^* = \left(T + \frac{1}{2} + \frac{1}{r}\right) \pm \sqrt{T^2 + T + \frac{1}{r^2} + \frac{1}{r} + \frac{1}{4}} \qquad (17)$$

Note that the case where k_0 and $\ln E_0$ are uncorrelated may not be tenable. However, this is not critical for the purpose of this study as long as the covariance is not too large and the bias in t^* is in the same direction for both males and females.

For clarification, note that if the covariance between k_0 and $\ln E_0$ is low, the second term on the RHS of (13) would determine the profile of log-earnings variance against experience level *t*. It is easy to see from (15) that the parabolic wage equation implies a fourth degree polynomial variance profile, where a minimum is reached around the early years of work experience, that is around t^*.

The intuition behind the U-shaped variance profile during these early years of work experience can be grasped by looking at equations (12) and (13), and considering two hypothetical individuals. First, ignore the last term on the RHS of (13). Assume that the two individuals are identical in every way except that person *H* invests more than person *L* in post-schooling human capital investment. That is $k_0^H > k_0^L$. From the first term on the RHS of (12), we can see that the intercept of the log-earnings profile of *H* would be lower than that of *L*. Furthermore, the third and fourth term on the RHS of (12) imply that the log-earnings profile for *H* will be steeper and more concave. The variation in log-earnings between the two individuals is minimised and is equal to zero when the two profiles cross at the experience level of *t**, which is termed the "overtaking experience level" by Mincer. After this experience level, the log-earnings profiles of the two individuals diverge again due to the difference in their initial net investment levels. The analysis in this paragraph is depicted diagrammatically in Figure 1.

FIGURE 1: Hypothetical Log-Earnings and Log-Earnings Variance Profiles



The analysis of this section extends straightforwardly to a sample of many individuals. The difference is that there is an additional variability from differences in the log of initial earnings potentials after schooling - $\ln E_0$ - in the sample, which acts as a vertical shifter of the variance profile.

RESEARCH METHODOLOGY

In the first stage, two separate equations of the form:

$$\ln Y_i = \hat{\beta}_0 + \hat{\beta}_t t_i + \hat{\beta}_{tsq} tsq_i + \hat{\beta}_s s_i + C_i'\hat{\beta} + \hat{u}_i$$

are estimated for both males and females, where *t* denotes the years of work experience, *tsq* denotes its square, *s* is the variable for schooling and *C* is a vector of control variables.

The next task is to determine the profile of the variance of the log-wage equation above with respect to the number of years of experience *t*, after relevant factors have been controlled for. Note that we need not calculate the variance at each level of experience explicitly. Since the variance at each t is approximately proportional to the squared of the OLS residuals $\hat{u}_i$ for individuals *i* at each *t*. Once a scatter plot of $\hat{u}_i^2$ against *t* is obtained, its smoothed profile can then be extracted by connecting the fitted values $(t_i, \hat{\hat{u}}_i^2)$ of the nonparametric regression through the scattered points. Each fitted point portrays the location of the distribution of $\hat{u}_i^2$ on each t_i. Specifically, the "locally weighted regression" technique with a "tricube" weight function is used (see Cleveland, 1979).

Once the variance minimising level of experience, t^* is determined, the next step is to obtain estimates of average marginal post-schooling return on investment *r*, the initial net investment ratio k_0 and the total investment horizon *T* for each gender. Recall that:

$$\beta_0 = \left(\ln Y_0 - k_0 - \frac{k_0^2}{2} \right) \qquad \text{(i)}$$

$$\beta_t = \left(rk_0 + \frac{k_0}{T} + \frac{k_0^2}{T} \right) \qquad \text{(ii)}$$

$$\beta_{tsq} = -\left(\frac{rk_0}{2T} + \frac{k_0^2}{2T^2} \right) \qquad \text{(iii)}$$

Using equations (ii) and (iii), the investment horizon *T* can be expressed as a function of k_0:

$$T(k_0) = \frac{-\hat{\beta}_t \pm \sqrt{\hat{\beta}_t^2 + 8\hat{\beta}_{tsq} k_0}}{4\hat{\beta}_{tsq}} \qquad (18)$$

Note that $\hat{\beta}_{tsq} < 0$ and that (18) yields sensible and non-negative values of *T* only when the numerator is negative. Using (17), the first expression for *r* is obtained:

$$r_1(k_0) = \frac{2(T(k_0) - t^*)}{2T(k_0)t^* - t^{*2} + t^*} \tag{19}$$

Note that t^* is the estimated variance minimising level of experience. In order to find a solution, a second expression for r is obtained from either equation (ii) or (iii). It makes no difference to the calculations whether (ii) or (iii) is used. Say, we use equation (ii):

$$r_2(k_0) = \frac{\hat{\beta}_t T(k_0) - k_0(1 + k_0)}{k_0 T(k_0)} \tag{20}$$

It can be shown that $r_1'(k_0)$ and $r_2'(k_0)$ approaches $-\infty$ when $T'(k_0)$ approaches $-\infty$. This occurs at the upper bound of k_0. Therefore the feasible region for k_0 is given by:

$$0 < k_0 < -\frac{\hat{\beta}_t^2}{8\hat{\beta}_{tsq}} \tag{21}$$

A solution will only exist if there is a k_0^* that lies within the region given in (21) that equates r_1 and r_2. Note that the existence of a solution is not guaranteed for the Mincer parabolic wage equation.

For convenience, and since equations (18)-(20) cannot be easily solved simultaneously, a graphical method is used to find a solution for r^*, k_0^* and T^* for the two genders. Once these estimates are found, $\ln Y_0^*$ can be solved directly from equation (i).

To accommodate the newly estimated information, the following modifications have to be made to the standard Oaxaca-Blinder decomposition. For each gender, consider the equation:

$$\overline{\ln Y} = \hat{\beta}_0 + \hat{\beta}_t \bar{t} + \hat{\beta}_{tsq} \overline{tsq} + \overline{X}\hat{\beta} \tag{22}$$

where the row vector $\overline{X}$ is composed of the mean years of schooling and the means of other relevant control variables. Using equations (i) to (iii), the first three terms on the RHS of (22) can be decomposed further and rearranged into $\overline{Z}b$ where:

$$\overline{Z} = \begin{bmatrix} 1 & k_0 & k_0^2 & k_0\bar{t} & \left(\frac{k_0 + k_0^2}{T}\right)\bar{t} & \frac{k_0}{T}\overline{tsq} & \frac{k_0^2}{T^2}\overline{tsq} \end{bmatrix} \tag{23}$$

$$b' = \begin{bmatrix} \ln Y_0 & -1 & -\frac{1}{2} & r & 1 & -\frac{r}{2} & -\frac{1}{2} \end{bmatrix} \tag{24}$$

Hence, for each group of male and female we have:

$$\overline{\ln Y} = \overline{Z}b + \overline{X}\hat{\beta} = \begin{bmatrix} \overline{Z} & \overline{X} \end{bmatrix} \begin{bmatrix} b \\ \hat{\beta} \end{bmatrix} \tag{25}$$

Since k_0 and T are the means of initial net investment in human capital ratio and the investment horizon respectively, they are to be considered average productive characteristics for each gender. On the other hand, $\ln Y_0$ can be interpreted as the mean of the initial log-wage before any formal human capital investment is undertaken (and after the effects of other control variables are factored out), and can be thought of as the market perception of the initial mean value of skills of individuals belonging to each gender group. Hence it is only appropriate that $\ln Y_0$ should be placed in one of the vectors of returns to characteristics, namely vector *b*.

All that remains to be done is just to calculate the new allocation of the average wage gap between male and female into the "explained" and "unexplained" portions using the following modified version of the Oaxaca-Blinder decomposition:

$$\overline{\ln Y_m} - \overline{\ln Y_f} = \begin{bmatrix} \left(\overline{Z}_m - \overline{Z}_f\right) & \left(\overline{X}_m - \overline{X}_f\right) \end{bmatrix} \begin{bmatrix} b_m \\ \hat{\beta}_m \end{bmatrix} + \begin{bmatrix} \overline{Z}_f & \overline{X}_f \end{bmatrix} \begin{bmatrix} (b_m - b_f) \\ (\hat{\beta}_m - \hat{\beta}_f) \end{bmatrix} \tag{26}$$

where the subscripts *m* and *f* denote male and female respectively. The new allocation is then compared to the original decomposition given in equation (3).

DATA ISSUES

This paper uses confidential data set from the Household, Income and Labour Dynamics in Australia (HILDA) survey. The HILDA project is managed by the Melbourne Institute of Applied Economic and Social Research (MIAESR) and is funded by the Commonwealth Department of Family and Community Services (FaCS). This longitudinal survey consists of four waves. This study uses data from the first wave only, which was completed in 2001 and involves 13,969 people.

The study is restricted to full-time workers - defined as those who work from 35 hours and above per week. The hourly earnings variable is calculated from dividing the reported current weekly gross wages and salary from all jobs by the combined hours per week currently worked in all jobs.

In constructing the years of schooling variable, individuals with the highest educational attainment of a master's degrees or higher are assigned 18 years of schooling. Those with graduate diplomas and those with bachelor's degrees as the highest levels are assigned 17 and 16 years respectively. Where a respondent holds an advanced diploma or a diploma, he/she is assigned 14 years. Those finishing year-12 and year-11 are assigned accordingly, and the remaining is classified as year-10-and-below. To avoid ambiguity, those individuals with certificates as the highest level of education achieved are dropped from the sample. The final sample consists of 2,188 males and 1,546 females.

Potential experience is used for the years of experience variable and is calculated by subtracting five years and the number of years of schooling from age. In addition to these key variables, our control variables that are thought to affect earnings include marital status, union membership, country of birth, region of residency and occupational and industry distribution. The summary statistics for these variables are shown in Table A1 in Appendix A.

EMPIRICAL RESULTS AND ANALYSIS

The first stage results from OLS regressions of the standard parabolic wage equations are reported in Table A2 in Appendix A for both males and females. To avoid biases resulting from differences in hours worked; the natural logarithms of hourly wage rates are used as the dependent variable. Also reported are the sample means for all the variables, as well as the robust standard errors and t-statistics for all the coefficients. Efforts were made to control for education, experience, personal, occupation, demographics, and industry of employment characteristics known to be important in determining the wage rate. Table 1 below highlights the three coefficients of interest for our study together with their formulae that were derived earlier.

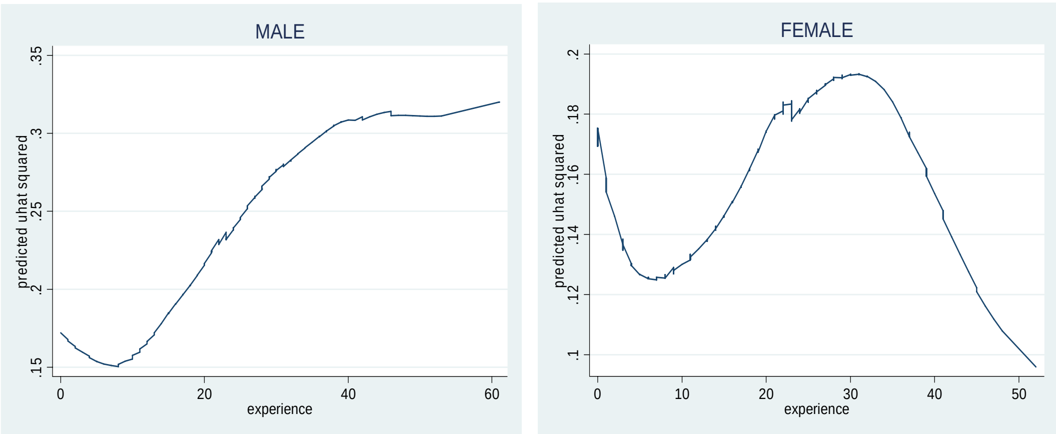
TABLE 1: OLS Estimates of Coefficients of Interest

		Male		**Female**	
Variable		Coefficient	t-Ratio	Coefficient	t-Ratio
Intercept	$\ln Y_0 - k_0 - k_0^2 / 2$	2.011	21.44	2.242	20.18
Experience	$rk_0 + k_0 / T + k_0^2 / T$	0.030	9.48	0.020	5.84
Experience Squared	$-rk_0 / 2T - k_0^2 / 2T^2$	-0.00054	-7.88	-0.00039	-5.12

A first glance at the three coefficients should cause us to suspect that there may be significant differences in the post-schooling investment in human capital between the genders. A relatively high intercept term of 2.242 for females relative to 2.011 for males, together with a lower steepness and smaller curvature of the female log-earnings-experience profile could be evidence of the trade-off mentioned above in the analyses of equation (12) and Figure 1.

In order to determine the log-earnings-variance minimising level of experience, $t*$, the squared OLS residuals from the first stage regression are scatter-plotted and a locally weighted scatter-plot smoothing is obtained by running a separate weighted regression on every data point. Points that are near the central point for each regression receive the highest weight. A tricube weight function with a bandwidth of 0.8 was chosen, meaning that 80% of the data are used in the smoothing of each data point. The predicted or smoothed points $(t_i, \hat{\hat{u}}_i^2)$ are then connected to obtain the log-earnings-variance profiles – "variance profile" for short - for males and females as are shown below in Figure 2.

FIGURE 2: Variance of Logarithm of Earnings - Experience Profiles



I focus on individuals in their early-to-mid careers since most data points are concentrated around this region and the wage structure, as well as the distribution of schooling should be most stable within this domain. From smoothing of the scatter-plots, a clear u-shaped variance profile emerges for each of the gender group. The variance minimising levels of experience $t*$ are estimated by fitting quadratic regressions around the troughs of the profiles and then using the estimated coefficients to calculate for t^*'s. These are estimated at 7.1 years and 6.6 years for males and females respectively. Analysis of Figure 1 alone is evidence that the average marginal returns to post-schooling human capital investment for males and females are similar, provided relevant factors are appropriately controlled for. Recall the analysis on equation (15) that if the investment horizon T is large, then t^* approaches $1/r$. Therefore, the first estimates of the average marginal returns to post-schooling investment for both genders are around 14%.

However, since we are equipped with equations (18) to (20), it is possible to go further and obtain more precise estimates for the average marginal return on investment in human capital r, the average investment horizon T and the average initial net investment ratio k_0 for each of the gender group. Once these estimates are obtained, $\ln Y_0$ which was interpreted above as the market perception of the initial mean value of skills of individuals belonging to each gender group can be calculated using equation (i). Graphical representation of the solution obtained using equations (18) to (20) is shown in Appendix B and the resulting estimates are summarised below in Table 2.

TABLE 2: Estimates of Unobserved Characteristics and Returns

Variable	**Male**	**Female**
t^* (years)	7.1	6.6
T^* (years)	19.30	16.67
$k_0{}^*$	0.18	0.11
r^*	10.6%	11.0%
$\ln Y_0{}^*$	2.21	2.36

It is clear from in Table 2 that the estimated average marginal returns to post-schooling investment (r) are very similar for our samples of males and females and are in the region of 11%. The biggest difference in productive characteristics is found in the estimated initial net human capital investment ratios (k_0). In other words, males on average initially invest around 18% of their productive time into augmenting their earnings potential in contrast to females who on average invest around 11% initially. The average investment horizon for males is estimated to be around 19 years, which is slightly longer than the 17 years average estimated for females.

With the newly estimated information the partial vector of mean characteristics $\overline{Z}$ and its corresponding partial coefficient vector b from equations (23) and (24) can now be constructed for our male and female samples. For exposition, the vectors are reproduced below:

$$\overline{Z} = \left[1 \quad k_0 \quad k_0^2 \quad k_0\bar{t} \quad \left(\frac{k_0 + k_0^2}{T}\right)\bar{t} \quad \frac{k_0}{T}\overline{tsq} \quad \frac{k_0^2}{T^2}\overline{tsq} \right]$$

$$b' = \left[\ln Y_0 \quad -1 \quad -\frac{1}{2} \quad r \quad 1 \quad -\frac{r}{2} \quad -\frac{1}{2} \right]$$

Note that $\bar{t}$ and $\overline{tsq}$ denote the sample means of the "experience" and "experience squared" variables respectively. The modified version of the Oaxaca-Blinder decomposition given in equation (26) that accommodates the newly estimated and unobserved returns and productive characteristics is also reproduced below:

$$\overline{\ln Y_m} - \overline{\ln Y_f} = \left[\left(\overline{Z}_m - \overline{Z}_f\right) \quad \left(\overline{X}_m - \overline{X}_f\right) \right] \begin{bmatrix} b_m \\ \hat{\beta}_m \end{bmatrix} + \left[\overline{Z}_f \quad \overline{X}_f \right] \begin{bmatrix} (b_m - b_f) \\ (\hat{\beta}_m - \hat{\beta}_f) \end{bmatrix}$$

The decomposition results using the original and the modified techniques applied to Australian 2001 data are presented together for ease of comparison below in Table 3. These results lend strong support to the suspicion that previous studies on gender wage differential based on the original Oaxaca-Blinder decomposition could have markedly underestimated the explained portion of the wage differential.

TABLE 3: Comparisons of the Two Versions of Oaxaca-Blinder Gender Wage Gap Decompositions for Australia, 2001

	Original Decomposition				**Modified Decomposition**			
Raw Wage Gap	Explained Gap	%	Unexplained Gap	%	Explained Gap	%	Unexplained Gap	%
0.138	0.016	11.4%	0.122	88.6%	0.065	46.8%	0.073	53.2%

The magnitude of the underestimation can be inferred from the above comparison using the same data set. Of the 13.8% difference in average hourly wage rate between the genders, the new decomposition finds that the explained portion accounts for around 47% of the difference, while using the original decomposition we would have been led to conclude that only 11% of the difference is attributable to differences in average observed characteristics.

SUMMARY AND CONCLUSION

The Oaxaca-Blinder decomposition has for more than thirty years served economists in the studies of earnings inequality. However, as is argued in this paper, this traditional method incompletely decomposes the difference in earnings between two comparative groups into a portion that is explained by differences in observed characteristics, and another that is attributed to differences in the valuation of these characteristics. Unless the average post-schooling human capital investment behaviours are the same across the groups under study, the traditional decomposition will underestimate the "explained" portion of the earnings gap.

To address the shortcoming of the Oaxaca-Blinder methodology, this study proposes for the first time an alternative decomposition technique that more completely decomposes the earnings gap. For illustration, the new decomposition is applied to the study of gender wage inequality in Australia using 2001 data from the HILDA survey. As is summarised above in Table 3, the explained portion of almost 47% of the hourly earnings gap calculated using the new decomposition is more than four-folds greater than that obtained using the traditional technique. In fact, this new estimate of the explained portion is much higher than estimates from all previous studies on gender wage inequality in Australia. As the results in Table 2 of the previous section indicate, compared to men, women on average invest less of their productive time into augmenting their human capital after formal schooling. In other words, a year of work experience for women should not be directly comparable to a year for their male counterparts. This is essentially what the new decomposition method takes into account.

There is a cost that comes with this greater insight, however; since we now require estimating the unobserved parameters of the post-schooling human capital investment function that are embedded within certain regression coefficients. Furthermore, and as mentioned above, given the assumption of constantly declining post-schooling human capital investment, the existence of a solution is not guaranteed. The linearly declining investment profile is itself not very realistic. We leave these areas for further research.

REFERENCES

BLINDER, A. S. (1973): “Wage Discrimination: Reduced Forms and Structural Estimates,” *Journal of Human Resources*, 8(4): 436-455.

CHAPMAN, B. J., and MULVEY, C. (1986): “An Analysis of the Origins of Sex Differences in Australian Wages,” *The Journal of Industrial Relations*, 28(4): 504-520.

CLEVELAND, W. S. (1979): “Robust Locally Weighted Regression and Smoothing Scatterplots,” *Journal of the American Statistical Association*, 74(368): 829-836.

CROCKETT, G. V. and PRESTON, A. C. (1999): “Effects of Labour Market Regulation on the Gender Pay Gap,” Curtin University of Technology, Curtin Business School, Women’s Economic Policy Analysis Unit (WEPAU), Discussion Paper Series 99/5.

HECKMAN, J. J., LOCHNER, L. J. and TODD, P. E. (2005): “Earnings Functions, Rates of Return, and Treatment Effects: The Mincer Equation and Beyond,” National Bureau of Economic Research, NBER working Paper Series, Working Paper 11544.

REIMAN, C. A. (1998): “Has Enterprise Bargaining Affected the Gender Wage Gap in Australia,” The University of Canberra, National Centre for Social and Economic Modelling (NATSEM), Discussion Paper, December.

MINCER, J. (1974): “Schooling, Experience and Earnings,” New York: National Bureau of Economic Research and Columbia University.

OAXACA , R. (1973): “Male-Female Wage Differentials in Urban Labour Markets,” *International Economic Review*, 14, 251-270.

APPENDIX A

TABLE A1: DESCRIPTIVE STATISTICS

	MALE					FEMALE				
Variable	**Obs**	**Mean**	**Std. Dev.**	**Min**	**Max**	**Obs**	**Mean**	**Std. Dev.**	**Min**	**Max**
Experience (years)	2188	20.43	11.71	0.00	61.00	1546	19.15	11.81	0.00	52.00
Schooling (years)	2188	13.09	2.96	6.00	18.00	1546	13.37	2.82	6.00	18.00
Age (years)	2188	38.52	11.44	15.00	72.00	1546	37.52	11.23	16.00	66.00
Hourly wage ($)	2188	21.83	13.04	0.28	132.22	1546	18.03	8.57	0.91	153.74
Log(hourly wage)	2188	2.94	0.56	-1.29	4.88	1546	2.80	0.46	-0.10	5.04
	Total	**Percent**				**Total**	**Percent**			
Total Observations	2188	100.00%				1546	100.00%			
Highest Schooling Attainment										
Postgraduate	137	6.26%				67	4.33%			
Graduate diploma	158	7.22%				145	9.38%			
Bachelor	451	20.61%				375	24.26%			
Advanced diploma	292	13.35%				203	13.13%			
Year 12	424	19.38%				314	20.31%			
Year 11	186	8.50%				111	7.18%			
Year 10 and below	540	24.68%				331	21.41%			
University graduates	746	34.10%				587	37.97%			
Non-university graduates	1442	65.90%				959	62.03%			
Marital Status										
Married	1565	71.53%				954	61.71%			
Single	623	28.47%				592	38.29%			
Union Membership										
Union	687	31.40%				523	33.83%			
Non-union	1501	68.60%				1023	66.17%			
Region										
Sydney	437	19.97%				322	20.83%			
Balance of NSW	240	10.97%				156	10.09%			
Melbourne	429	19.61%				329	21.28%			
Balance of Victoria	138	6.31%				83	5.37%			
Brisbane	210	9.60%				160	10.35%			
Balance of QLD	207	9.46%				145	9.38%			
Adelaide	132	6.03%				70	4.53%			
Balance of SA	44	2.01%				34	2.20%			
Perth	169	7.72%				124	8.02%			
Balance of WA	53	2.42%				33	2.13%			
Tasmania	56	2.56%				43	2.78%			
Northern Territory	11	0.50%				15	0.97%			
ACT	62	2.83%				32	2.07%			
Occupation										
Managers and Administrators	290	13.25%				110	7.12%			
Professionals	577	26.37%				520	33.64%			
Associate Professionals	291	13.30%				228	14.75%			
Tradespersons and related workers	230	10.51%				37	2.39%			
Advanced Clerical and service workers	17	0.78%				109	7.05%			
Intermediate Clerical, Sales and Service Workers	209	9.55%				334	21.60%			
Intermediate Production and Transport Workers	299	13.67%				36	2.33%			
Elementary Clerical, Sales and Service Workers	89	4.07%				100	6.47%			
Labourers and Related Workers	186	8.50%				72	4.66%			

	MALE		FEMALE	
	Total	**Percent**	**Total**	**Percent**
Agriculture, Forestry and Fishing	116	5.30%	24	1.55%
Mining, Oil and Gas Extraction	48	2.19%	6	0.39%
Manufacturing	330	15.08%	124	8.02%
Electricity, Gas and Water	31	1.42%	7	0.45%
Construction and Construction Trade Services	163	7.45%	23	1.49%
Wholsaling	133	6.08%	51	3.30%
Retailing	171	7.82%	146	9.44%
Restaurants and Cafes	56	2.56%	57	3.69%
Transport and Storage	143	6.54%	33	2.13%
Communication	79	3.61%	29	1.88%
Finance and Insurance	109	4.98%	97	6.27%
Property and Business Services	296	13.53%	219	14.17%
Government Administration and Defense	133	6.08%	95	6.14%
Education	151	6.90%	256	16.56%
Health and Community Services	100	4.57%	290	18.76%
Recreation and Other Services	129	5.90%	89	5.76%
Country of Birth				
Australia	1608	73.49%	1158	74.90%
Main English Speaking	266	12.16%	172	11.13%
Other	314	14.35%	216	13.97%

TABLE A2: OLS REGRESSION RESULTS

Dependent Variable: Natural Logarithm of Hourly Wage

	Male				Female			
Variable	Coefficient	Robust S.E.	t-Ratio	Mean of X	Coefficient	Robust S.E.	t-Ratio	Mean of X
Intercept	2.011	0.094	21.44	1	2.242	0.111	20.18	1
Schooling	0.050	0.005	9.75	13.093	0.022	0.006	3.92	13.373
Experience	0.030	0.003	9.48	20.431	0.020	0.003	5.84	19.151
Experience Squared	-0.00054	0.000	-7.88	554.511	-0.00039	0.000	-5.12	506.100
Married	0.068	0.022	3.05	0.715	-0.001	0.021	-0.06	0.617
Union Membership	0.126	0.021	6.09	0.314	0.125	0.023	5.41	0.338
Region of Residency[1]								
Sydney	-0.055	0.053	-1.04	0.200	0.087	0.059	1.48	0.208
Balance of NSW	-0.225	0.059	-3.82	0.110	-0.064	0.061	-1.06	0.101
Melbourne	-0.068	0.052	-1.30	0.196	0.032	0.059	0.55	0.213
Balance of Victoria	-0.237	0.060	-3.95	0.063	-0.133	0.081	-1.64	0.054
Brisbane	-0.200	0.062	-3.26	0.096	-0.025	0.065	-0.38	0.103
Balance of QLD	-0.227	0.059	-3.86	0.095	-0.147	0.065	-2.25	0.094
Adelaide	-0.211	0.062	-3.42	0.060	-0.092	0.070	-1.33	0.045
Balance of SA	-0.283	0.098	-2.91	0.020	-0.258	0.110	-2.35	0.022
Perth	-0.177	0.060	-2.94	0.077	0.007	0.062	0.12	0.080
Balance of WA	-0.031	0.082	-0.38	0.024	-0.050	0.081	-0.61	0.021
Tasmania	-0.215	0.074	-2.93	0.026	-0.106	0.090	-1.18	0.028
Northern Territory	0.046	0.101	0.46	0.005	0.033	0.069	0.49	0.010
Country of Birth[2]								
English-Speaking	0.021	0.033	0.64	0.122	-0.011	0.038	-0.30	0.111
Others	-0.083	0.029	-2.83	0.144	-0.034	0.028	-1.23	0.140
Occupational Distribution[3]								
Managers and Administrators	0.151	0.049	3.08	0.133	0.438	0.068	6.42	0.071
Professionals	0.151	0.046	3.27	0.264	0.334	0.055	6.08	0.336
Associate Professionals	0.090	0.046	1.97	0.133	0.162	0.054	3.01	0.147
Tradesperson and Related Workers	-0.023	0.045	-0.51	0.105	-0.042	0.096	-0.44	0.024
Advanced Clerical and Service Workers	0.010	0.107	0.09	0.008	0.165	0.059	2.77	0.071
Intermediate Clerical, Sales and Services	0.027	0.042	0.65	0.096	0.079	0.047	1.69	0.216
Intermediate Production and Transport	-0.002	0.037	-0.05	0.137	-0.081	0.084	-0.97	0.023
Elementary Clerical and Service Workers	-0.109	0.054	-2.01	0.041	0.070	0.066	1.06	0.065
Industry Distribution[4]								
Agriculture, Forestry and Fishing	-0.265	0.064	-4.15	0.053	-0.253	0.145	-1.74	0.016
Mining, Oil and Gas Extraction	0.353	0.066	5.34	0.022	0.154	0.116	1.33	0.004
Manufacturing	-0.013	0.035	-0.36	0.151	-0.156	0.047	-3.35	0.080
Electricity, Gas and Water	0.127	0.052	2.43	0.014	-0.118	0.114	-1.04	0.005
Construction and Construction Trade Services	0.014	0.042	0.33	0.074	-0.075	0.060	-1.25	0.015
Wholesaling	-0.120	0.047	-2.56	0.061	-0.111	0.071	-1.56	0.033
Retailing	-0.214	0.045	-4.74	0.078	-0.283	0.052	-5.48	0.094
Cafes and Restaurants	-0.256	0.065	-3.91	0.026	-0.203	0.071	-2.88	0.037
Transport and Storage	-0.057	0.051	-1.13	0.065	-0.007	0.071	-0.10	0.021
Communication	0.083	0.046	1.79	0.036	0.122	0.089	1.37	0.019
Finance and Insurance	0.235	0.053	4.40	0.050	-0.069	0.045	-1.52	0.063
Property and Business Services	0.027	0.043	0.63	0.135	-0.047	0.044	-1.07	0.142
Education	-0.215	0.039	-5.58	0.069	-0.196	0.042	-4.71	0.166
Health and Community Services	-0.044	0.054	-0.81	0.046	-0.160	0.042	-3.84	0.188
Recreation and Other Services	-0.237	0.053	-4.47	0.059	-0.225	0.062	-3.62	0.058
R-squared		0.354				0.316		
Adjusted R-Squared		0.341				0.296		
Mean of ln(Y)		2.936				2.798		
Number of Observations		2188				1546		

Note: 1. Reference group ACT, 2. Reference group Australian, 3. Reference group Labourers and Related Workers, 4. Reference group Government Administration and Defense

APPENDIX B

FIGURE B1: Intersection of the Two Rate of Return to Post-Schooling Human Capital Investment Functions for Males and Females

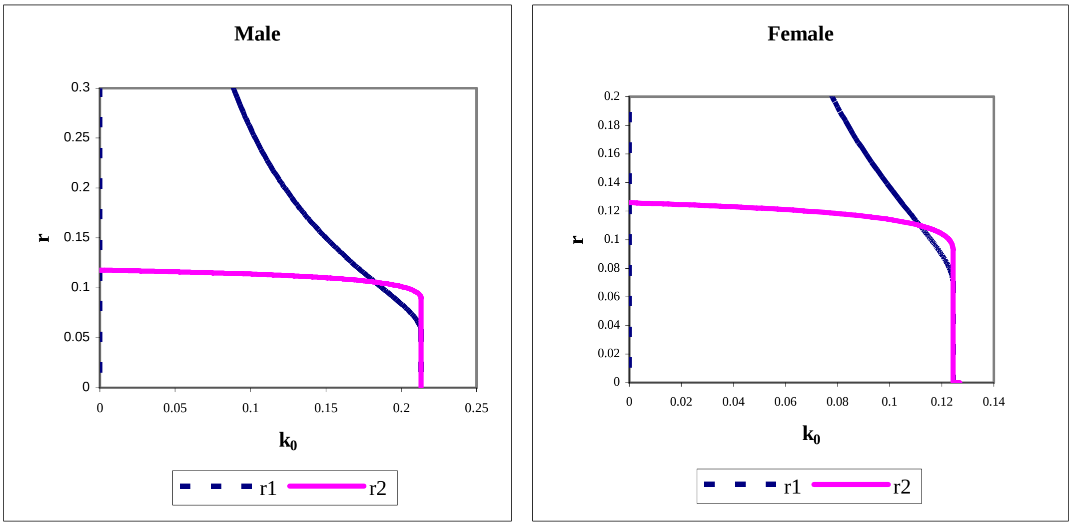


FIGURE B2: Post-Schooling Investment Horizon as a Function of Initial Net Investment Ratios for Males and Females