H.Ü. İktisadi ve İdari Bilimler Fakültesi Dergisi, Cilt 18, Sayı 2, 2000, s. 283-303

# MALE-FEMALE EARNINGS DIFFERENTIALS AND THE EXTENT OF WAGE DISCRIMINATION IN PROFESSIONAL OCCUPATIONS†

**Burak GÜNALP**

*(Asst.Prof., Department of Economics, Hacettepe University TR-06532, ANKARA e-mail: gunalp@hacettepe.edu.tr)*

*(Yrd.Doç.Dr., Hacettepe Üniversitesi, İktisat Bölümü)*

**Abstract:**

Using data from a National Science Foundation survey, this study measures the extent of gender discrimination in the case of college graduates. To analyze the components of the male-female earnings gap, the standard Oaxaca/Blinder decomposition method and the Cotton/Neumark extensions of this method are employed. In addition, the study accounts for the sample selection bias by using Heckman's two-stage, two-equation procedure. The results of the study indicate that a significant portion of the observed earnings differential is attributable to wage discrimination.

**Özet:**

**Profesyonel Mesleklerde Kadın-Erkek Ücret Farklılıkları ve Ücret Ayırımcılığının Boyutları**

Bu çalışmada, Ulusal Bilim Vakfı verileri kullanılarak, üniversite mezunları için, kadın-erkek ücret ayırımcılığının hangi boyutlarda olduğu ortaya konulmaktadır. Ücret farklılıklarının bileşenlerinin analiz edilmesinde, Oaxaca ve Blinder tarafından geliştirilen standart ayrıştırma yöntemi ve bu yöntemin Cotton ve Neumark tarafından geliştirilen uzantıları kullanılmaktadır. Ayrıca, örneklem seçim problemi için Heckman'ın iki

---

† This study is a revised version of the paper presented at METU International Conference on Economics, September 9-12, 1998. The helpful comments and suggestions from the participants in this conference are gratefully acknowledged.

**Keywords:** Labor market discrimination, male-female earnings differentials, male-female wage discrimination, gender discrimination.

**Anahtar Sözcükler:** Emek piyasası ayırımcılığı, kadın-erkek ücret farklılıkları, kadın-erkek ücret ayırımcılığı, cinsiyet ayırımcılığı.

aşamalı, iki denklemli yöntemi kullanılmaktadır. Çalışmanın sonuçları, gözlemlenen ücret farkının önemli bir kısmının ücret ayırımcılığına atfedilebileceğini göstermektedir.

## I. INTRODUCTION

One of the most important developments in the U.S. labor market since World War II has been the increase in the labor force participation of women. This has resulted from a reduction in the reservation wage of women and from the rising real wages that women can obtain in the labor market. Throughout this period an increase in the ratio of female to male earnings has been observed. However, there is still a substantial earnings differential in favor of men.

A sex differential in wages is considered to be discriminatory if the differential cannot be explained by sex differences in productivity. Numerous studies have been performed to measure the extent of gender discrimination in the U.S. labor markets. Most of these studies have found a substantial amount of earnings differential after adjusting for sex differences in productivity. This differential has been attributed to labor market discrimination against women.[1]

Using data from a 1986 National Science Foundation survey, this paper aims to measure the extent of gender discrimination in the case of college graduates (more specifically, scientists and engineers). As indicated above, there are countless studies on gender discrimination in the U.S. labor markets. However, just a few of these concentrate on highly educated professional employees.[2] A study of sex discrimination in a graduate labor market may be motivated in two ways. First, graduate labor markets are important for the U.S. economy. Since a large majority of professional workers are graduates, it is important to analyze whether their wages are unfairly conditioned by sex. Second, using a sample restricted to particular job categories has some advantages. The most important reason for analyzing such a sample is a desire for homogeneity. Restricting analysis to specific groups of employees reduces substantially the range of variation in skill, education, and other forms of training. This results in a considerably more homogeneous set of individuals.[3]

The study consists of five sections. In Section II we describe the methods used to measure discrimination and briefly explain Heckman's two-stage, two-equation procedure for correcting the problem of sample selection bias. The empirical specifications are outlined in Section III. In Section IV we present the estimation results and calculate the components of the male-female earnings gap. The last section is the concluding section.

## II. THE METHODOLOGY

The standard method to analyze the components of the male-female earnings differential is the decomposition approach proposed by Oaxaca (1973) and Blinder (1973). This method requires the estimation of earning equations for samples of individual men and women separately. Then, using these estimates the overall average earnings differential is decomposed into two components: One is the portion attributable to differences in the endowments of wage-generating characteristics, and the other is taken as reflecting discrimination.

As mentioned above the method requires first estimating Equation (1) to a sample of male (m) workers and Equation (2) to a sample of female (f) workers:[4]

$$W_m = \beta_m X_m + u_m \qquad (1)$$

$$W_f = \beta_f X_f + u_f \qquad (2)$$

where $W$ is the natural logarithm of wages or earnings, $X$ is a vector of productivity-enhancing characteristics of the workers (education, experience, etc.) and several control variables (marital status, race, location, etc.), $\beta$ is the vector of regression coefficients, and $u$ is a random error term.

Ordinary least-squares regression analysis has the property that the fitted regression lines always pass through the point of sample means. This implies that

$$\overline{W_m} = b_m \overline{X_m} \qquad (3)$$

$$\overline{W_f} = b_f \overline{X_f}. \qquad (4)$$

Here $b$'s are the least-squares estimates of $\beta$'s. If females obtain the same return as do males for their endowments of wage determining characteristics (in other words, if females were given the male pay structure), then their average wage would be

$$\overline{W_f^*} = b_m \overline{X_f}. \qquad (5)$$

This can be interpreted as the average female wage that would be valid if there were no wage discrimination. Here wage discrimination is defined as unequal pay for the same endowments of wage-determining characteristics.

Subtracting Equation (5) from Equation (3) yields the difference between average male earnings and the average hypothetical female earnings that would be received by females if they were paid according to the male pay structure. In other words, this difference represents the different endowments of wage-generating characteristics of the two groups, that is,

$$\overline{W_m} - \overline{W_f^*} = b_m \overline{X_m} - b_m \overline{X_f} = b_m (\overline{X_m} - \overline{X_f}). \tag{6}$$

Subtracting Equation (4) from Equation (5) gives the difference between the hypothetical *nondiscriminatory* wage and the actual wage of females. This difference represents the different returns to the same wage-generating characteristics, that is,

$$\overline{W_f^*} - \overline{W_f} = b_m \overline{X_f} - b_f \overline{X_f} = (b_m - b_f) \overline{X_f}. \tag{7}$$

Finally, adding Equation (6) and Equation (7) yields

$$\overline{W_m} - \overline{W_f} = b_m (\overline{X_m} - \overline{X_f}) + (b_m - b_f) \overline{X_f}. \tag{8}$$

That is, we decompose the overall average male-female earnings differential into two components: $b_m (\overline{X_m} - \overline{X_f})$ and $(b_m - b_f) \overline{X_f}$. The first component is attributable to differences in the endowments of wage-generating characteristics of the two groups $(\overline{X_m} - \overline{X_f})$ evaluated at the male returns $(b_m)$. The second component, on the other hand, is attributable to differences in the returns $(b_m - b_f)$ that males and females get for the same endowment of wage-generating characteristics $(\overline{X_f})$. This second component is taken as a measure for the extent of wage discrimination.

An alternative decomposition of the earnings differential can be derived by replacing Equation (5) with the hypothetical wage that males could expect to earn if they were paid according to the female pay structure:

$$\overline{W_m} - \overline{W_f} = b_f (\overline{X_m} - \overline{X_f}) + (b_m - b_f) \overline{X_m}. \tag{9}$$

The difference between Equation (8) and Equation (9) is that in the first one we assume that the current male wage structure would apply to both males and females in the absence of discrimination. And, in the second one we assume that current female wage structure would apply to both males and females in a nondiscriminating labor market. The two equations are alternative representations of the decomposition and neither is preferable to the other a priori. Even though the two equations will not produce equivalent results, there is no theoretical reason or advantage to choose one equation over the other. This gives rise to the so-called index number problem. Since economic theory does not provide much guidance on this, many studies report both of the estimates.

A third alternative is suggested by Cotton (1988). Cotton starts with Becker's (1957) assumption that in the absence of discrimination in perfectly competitive markets males and females would be perfect substitutes in production. In other words, in the absence of discrimination wage differences would result solely from productivity characteristics. Hence, in the absence of discrimination the wage structures are the same and equal to $\beta^*$ for both males and females. Cotton decomposes the average male-female earnings differential as:

$$\overline{W_m} - \overline{W_f} = b^*(\overline{X_m} - \overline{X_f}) + (b_m - b^*)\overline{X_m} + (b^* - b_f)\overline{X_f}, \qquad (10)$$

where $b^*$ is the estimated nondiscriminatory wage structure. The first term on the right hand side of the equation is an estimate of the productivity differential. The discrimination component is made up of two elements in this decomposition: the second and the third terms on the right hand side of the equation. The second term represents the amount by which male productivity characteristics are overvalued (the benefit of being a male worker). The third term represents the amount by which female productivity characteristics are undervalued (the cost of being a female worker). An empirical problem here is how to estimate the unobservable nondiscriminatory wage structure $\beta^*$. Cotton assumes that the wage structure that would exist in the absence of discrimination is the simple weighted average of the observed structures for males and females, weights being the proportions of males and females in the employed labor force (or the whole regression sample).[5] That is, $b^* = l_m b_m + l_f b_f$, where $l_m$ and $l_f$ are the proportions.

A useful representation of the estimated nondiscriminatory wage structure is given by

$$b^* = \Omega b_m + (I - \Omega) b_f, \tag{11}$$

where $\Omega$ is a weighting matrix. As explained by Oaxaca and Ransom (1994), given $b_m, b_f$ and Equation (11), any assumption about $b^*$ reduces to an assumption about $\Omega$. For example, Oaxaca (1973) proposes that either the current male wage structure (Equation (8)) or the current female wage structure (Equation (9)) should be adopted as the nondiscriminatory wage structure. Choosing the weighting matrix $\Omega = I$ and substituting Equation (11) in Equation (10) will yield Equation (8) while choosing the weighting matrix $\Omega = 0$ and substituting Equation (11) in Equation (10) will yield Equation (9). Cotton (1988) in effect chooses the weighting matrix $\Omega = l_m I$ where $l_m$ is the proportion of male workers in the sample as mentioned above.

Neumark (1988) argues that the nondiscriminatory wage structure should be derived from a theoretical model of discriminatory behavior and accomplishes this by extending Becker's (1957) and Arrow's (1973) model of discrimination model. According to Neumark, the nondiscriminatory wage structure and the appropriate decomposition depend on the nature of employers' discriminatory behavior. Employers may practice nepotism toward males or discrimination against females. Under nepotism toward males (and no discrimination against females), females are paid the competitive wage, but males are overpaid. In such a situation, it is appropriate to use the female wage structure, $b_f$, as the nondiscriminatory wage structure. Under discrimination against females (and no nepotism toward males), males are paid the competitive wage, but females are underpaid. In such a situation, it is appropriate to use the male wage structure, $b_m$, as the nondiscriminatory wage structure. Neumark assumes that, in reality employers can be both nepotistic toward males and discriminatory against females. The cost of relaxing the pure nepotism or pure discrimination assumption is that a restriction must be imposed on employers' preferences to be able to derive an estimable nondiscriminatory wage structure. The restriction imposed is that, within each type of labor, the utility function is homogenous of degree zero with respect to male and female labor inputs, that is to say, employers care only about the relative proportions of males and females and not absolute numbers. With such a restriction, Neumark shows that the nondiscriminatory wage structure can be estimated from an earnings function estimated over the pooled sample of males and females. That is:

$$b^* = (X'X)^{-1}(X'Y) = b, \tag{12}$$

where $b$ is the OLS estimate obtained from the pooled sample of males and females. Oaxaca and Ransom (1994) shows that the weighting matrix, in this case, is $\Omega = (X'X)^{-1}(X'_m X_m)$.

In this paper we will decompose the male-female earnings differential in our sample according to the four decomposition methods explained above: The methods proposed by Oaxaca (1973) and Blinder (1973), that is, the methods in which the nondiscriminatory wage structure is the male wage structure (Equation (8)) and the nondiscriminatory wage structure is the female wage structure (Equation (9)); the method proposed by Cotton (1988) in which the the nondiscriminatory wage structure is the weighted average of the wage structures for males and females; and the method proposed by Neumark (1988) in which the nondiscriminatory wage structure is calculated from an earnings function estimated over the pooled sample of males and females.[6]

A problem encountered in estimating earnings equations is the sample selection bias arising from the fact that earnings samples include only working women and men since earnings data are available only for them. One of the assumptions of the classical linear regression model is that the expected value of the random error term in the sample is zero. However, for certain individuals the wages earned were they to participate in the labor market would not be sufficient, and these individuals therefore choose not to be in the labor force. In other words, reservation wages of such individuals are greater than the market wage. Since these people do not work, their wages are not observed in the sample. Thus earnings samples include only working women and men since earnings data are available only for them. Under plausible assumptions it can be shown that when some individuals are selectively excluded from the sample, the expected value of the error term is likely to be nonzero for the observed sample although it may be zero for the total population. As a result, estimating earnings equations by ordinary least squares using a sample restricted only to workers may result in biased estimates of the intercept and slope coefficients. This resulting bias is called the sample selection bias.

The most widely used method for analyzing nonrandomly selected samples is the two-stage, two-equation procedure developed by Heckman (1979), which yields consistent estimates of the parameters. Heckman shows that the sample selection problem can be viewed as a specification error in which a relevant variable is incorrectly omitted from the regression equation. To correct this problem, Heckman suggests a two-step procedure: (1) Using data from all workers and nonworkers, estimate the inverse of Mill's ratio $(\lambda)$ from a probit equation of the probability of an individual being in the labor force (i.e., the probability of an individual being observed in the wage regression sample); (2) Include $\lambda$ as a regressor in the regression equation (earnings equation in our study) and estimate this equation by OLS. Heckman shows that this procedure yields consistent estimates of the parameters.[7] To account for the sample selection problem, we will use Heckman's two-stage, two-equation procedure.

## III. THE DATA AND THE EMPIRICAL SPECIFICATION

The data used in the study come from *1986 Survey of Scientists and Engineers* conducted by the Bureau of the Census for the National Science Foundation. There are 29 categories in the occupational classification of the survey. This survey is well suited to the issue raised in this paper because detailed information on earnings, work experience, education, location of residence, etc. was collected for a large sample of men and women. The resulted sample contains 45752 individuals of which 38899 are men and 6853 are women. These numbers are appropriate to represent total male and female populations in the occupations examined.

The empirical specification of earnings equations (Equations (1) and (2)) is based on the standard human capital earnings function developed by Mincer (1974) which has guided much of the empirical research into the determinants of earnings:[8]

$$\begin{aligned} W = {} & \beta_0 + \beta_1 EDUC + \beta_2 EXP + \beta_3 EXPSQ + \beta_4 TENURE + \\ & \beta_5 WKSWK + \beta_6 WHITE + \beta_7 PRIVEM + \beta_8 GOVEM + \\ & \beta_9 SELFEM + \beta_{10} US + \beta_{11} MARRIED + \beta_{12} NORTH + \\ & \beta_{13} SOUTH + \beta_{14} MOUNT + \beta_{15} PACIFIC + \beta_{16} \lambda + u. \end{aligned} \tag{13}$$

Variables are listed and described in Table 1. An examination of the mean values of the variables reveals that, on average, women earn 62.9% of men's monthly earnings but work fewer weeks per year. In addition, women have, on average, substantially less experience than male workers. A similar structure is observed also for *TENURE* variable. With respect to education (*EDUC*), on the other hand, women average slightly higher than men. Also, marriage ratio is higher for males than for females.[9]

As explained in Section II, this study accounts for the sample selection bias by using Heckman's (1979) two-stage, two-equation procedure. According to this procedure, in the next section we will first calculate the inverse of Mill's ratio,

$(\lambda)$, from probit estimates for the probability of labor force participation using data from all workers and nonworkers. Then we will include $\lambda$ (for workers only) as an independent variable in the earnings regression equations. The independent variables included in our labor force participation equation are standard variables employed in many studies estimating such an equation.[10] Our labor force participation equation contains thirteen variables. Eight of these variables are the same as those used in our earnings equation:[11] *EDUC*, *WHITE*, *US*, *MARRIED*, *NORTH*, *SOUTH*, *MOUNT* and *PACIFIC*. Their descriptions are given in Table 1. In addition to these eight variables, the labor force participation equation includes five new variables: *AGE*, *AGESQ*, *CHOLD*, *CHYOUNG* and *STUDENT*. These variables are defined as follows:

| | | |
|---|---|---|
| *AGE* | : | Age of the person. |
| *AGESQ* | : | Age squared. |
| *CHOLD* | : | 1 if there are 6-17 year-old children living with the person, 0 otherwise. |
| *CHYOUNG* | : | 1 if there are children 5 years old and under living with the person, 0 otherwise. |
| *STUDENT* | : | 1 if the person is currently attending a college or university, 0 otherwise. |

The dependent variable in the probit equation is a binary variable defined to equal one if the person is employed at the time of the survey, zero otherwise. The results of estimating the labor force participation equation and the earnings equations are presented below.

**Table 1:** Variables

| Abbreviated Name | Description |
|---|---|
| *W* | The natural logarithm of monthly income. (The dependent variable.) |
| *EDUC* | Years of formal education completed beyond high school. |
| *EXP* | Years of professional work experience. |
| *EXPSQ* | Experience squared. |
| *TENURE* | Length of service in the last firm. |
| *WKSWK* | Weeks worked. |
| *WHITE* | 1 if the person is white, 0 otherwise. (The reference group is "nonwhite".) |
| *PRIVEM* | 1 if the person is a private employee, 0 otherwise. |
| *GOVEM* | 1 if the person is a government employee, 0 otherwise. |
| *SELFEM* | 1 if the person is self-employed, 0 otherwise. (The reference group for the last three variables is "other" including 'no response' and 'unpaid family worker' categories.) |
| *US* | 1 if the person is a U.S. citizen, 0 otherwise. (The reference group is "non-U.S. citizen".) |
| *MARRIED* | 1 if the person is married, 0 otherwise. (The reference group is "unmarried".) |
| *NORTH* | 1 if the person's location of residence is at the northern part of the U.S., 0 otherwise. |
| *SOUTH* | 1 if the person's location of residence is at the southern part of the U.S., 0 otherwise. |
| *MOUNT* | 1 if the person is from 'mountain' region, 0 otherwise. |
| *PACIFIC* | 1 if the person is from 'pacific' region, 0 otherwise. (The reference group for the last four variables is "other" including 'all outlying U.S. territories' and 'all foreign countries'.) |
| $\lambda$ | The inverse of Mill's ratio. |

## IV. ESTIMATION RESULTS

The results of the probit equation for the probability of labor force participation for men and women are given in Table 2.[12] It is seen from this table that for male workers there is a significant negative association between educational attainment and the likelihood of being in the labor force. For female workers, on the other hand, *EDUC* is not a significant determinant of participation. The negative and significant coefficient on *EDUC* for male workers is not consistent with our expectations. However, it might be true for men that an individual with higher education would have a higher reservation wage because of raised expectations and would be less willing to take a lower paying job than someone with low education.

It is also seen that for both men and women being white and being a U.S. citizen seem to be significant determinants of labor force participation. Positive sign on the coefficient of *WHITE* indicates that white individuals are more likely to be in the labor force. This may suggest the fact that higher wages are paid to whites. Besides, it is also true that the likelihood of finding a job is greater for white individuals. Highly significant and negative coefficient on *US* variable is consistent with the findings of many other studies. This may reflect the facts that the U.S. citizens have usually higher reservation wages than foreigners and that foreigners are, in general, better educated than the U.S. citizens.

The coefficient of *MARRIED* is positive and highly significant for men while it is negative but insignificant for women. The positive coefficient for male workers may indicate that a married man with responsibilities for his family has a lower reservation wage compared to a single individual.

We expect that a woman who has young children would be less likely to work than a woman who has no children because her time in the household is more valuable to her. The result is consistent with this expectation. The coefficient of *CHYOUNG* variable is negative and significant at the 5 percent level for women.

One would also expect that individuals who are presently students are less likely to participate in the labor force than nonstudents. Unexpectedly, however, the coefficient of *STUDENT* is positive and significant at the 1 percent level for both men and women. This may be true if the students in the sample are, in general, working persons whose education expenses are covered by their employers.

The human capital theory of labor supply suggests that the pattern of labor force participation rates by age is shaped like an inverted U. For a young and inexperienced individual the market wage is probably low, providing little incentive to participate in the labor force. Besides, young individuals are more likely to go to school and postpone labor-force entry in order to accumulate human capital. Schooling is one form of accumulating human capital. An individual can accumulate human capital also by on-the-job training. As the individual matures and accumulates human capital through on-the-job training, his or her market wage will increase, implying a greater likelihood of labor force participation. Finally, as the individual gets older, his or her incentives for investing in on-the-job training will decline, and eventually, the human capital accumulated will start to depreciate, causing the individual's market wage to decrease. Furthermore, the value of leisure time and thus the reservation wage will be greater for an older individual because of the wealth he or she has accumulated during his or her working life. As a result, an older individual will be less likely to be in the labor market.

**Table 2:** Probit Estimates for Male and Female Labor Force Participation

| Variable | Male | | Female | |
|---|---|---|---|---|
| | **Coefficients** | **Asymptotic t-ratios** | **Coefficients** | **Asymptotic t-ratios** |
| *Constant* | 1.969 | 11.316 | 1.736* | 5.425 |
| *EDUC* | -0.034* | -9.444 | 0.0083 | 1.038 |
| *WHITE* | 0.245* | 10.000 | 0.116** | 2.320 |
| *US* | -1.202* | -46.230 | -0.980* | -16.07 |
| *AGE* | 0.104* | 19.259 | 0.099* | 8.250 |
| *AGESQ* | -0.0014* | -23.729 | -0.0013* | -9.220 |
| *MARRIED* | 0.378* | 17.260 | -0.0243 | -0.639 |
| *NORTH* | 0.056 | 0.448 | -0.250 | -1.214 |
| *SOUTH* | 0.041 | 0.328 | -0.297*** | -1.435 |
| *MOUNT* | 0.072 | 0.563 | -0.257 | -1.190 |
| *PACIFIC* | -0.032 | -0.254 | -0.335*** | -1.610 |
| *CHOLD* | -0.093* | -4.895 | -0.057 | -1.188 |
| *CHYOUNG* | -0.020 | -0.870 | -0.387** | -2.413 |
| *STUDENT* | 0.112* | 3.733 | 0.210* | 4.468 |
| Log Likelihood | -16417.97 | | -3437.42 | |
| Sample Size | 42906 | | 8412 | |

Note: *indicates statistical significance at the 1 percent level.
**indicates statistical significance at the 5 percent level.
***indicates statistical significance at the 10 percent level.

The human capital theory of labor supply, therefore, implies that the probability of labor force participation should be quadratic in age; that is, the labor force participation equation should contain both *AGE* and *AGESQ* variables, with a positive sign expected for the coefficient of *AGE* and a negative sign for the coefficient of *AGESQ*.

For both men and women the estimated coefficients of AGE and AGESQ variables are positive and negative, respectively, as expected, and highly significant. This is consistent with the predictions of the human capital theory of labor supply.

Finally, the coefficients of regional dummies (*NORTH, SOUTH, MOUNT* and *PACIFIC*) are, in general, very insignificant indicating that being from a specific region is not an important determinant of participation.

As we explained in the previous section, we calculate the inverse of Mill's ratio ($\lambda$) from these probit estimates and include it as an independent variable in the earnings equations.

The results of estimating the earnings equations are given in Table 3. It is seen that for both models the coefficients of multiple determination ($R^2$) are low but they are still acceptable for a cross-section study of wage determination, and the F statistics are highly significant. Thus the overall performance of each model seems quite good. Looking at the individual coefficients, we observe that most of the coefficients are significant at the 1 percent level, and their signs are mostly as expected.

It is also seen from Table 3 that the coefficient of the inverse of Mill's ratio ($\lambda$) is statistically significant and negative for both men and women, suggesting that individuals who were employed at the time of the survey are not a random sample of all individuals. This suggests that if we had estimated our earnings equations without accounting for selectivity bias, the parameter estimates could have been biased. Since the inverse of Mill's ratio is a monotonic decreasing function of the probability that an individual is selected into the sample, a negative coefficient on $\lambda$ implies that higher earnings are associated with a greater probability of participation (Heckman, 1979; Dolton and Makepeace, 1987).

The coefficient of *EDUC* is positive for both men and women as expected but it is significant only for men. Besides, the value of this coefficient is substantially greater for men than for women indicating that men have a greater rate of return to education compared to women. A positive coefficient on this variable implies that individuals with higher levels of education earn more in the market. As can be seen from the table, the estimated rate of return to schooling for men is 0.025. That is, one-year increase in educational attainment leads to 2.5 percent increase in monthly earnings.

**Table 3:** OLS Estimates of Earnings Equations
(Dependent Variable: Natural Logarithm of Monthly Income)

| Variable | Male | | Female | |
|---|---|---|---|---|
| | Coefficients | Asymptotic t-ratios | Coefficients | Asymptotic t-ratios |
| *Constant* | 6.836* | 156.50 | 6.81* | 59.05 |
| *EDUC* | 0.025* | 24.32 | 0.0034 | 1.166 |
| *EXP* | 0.035* | 40.48 | 0.036* | 15.12 |
| *EXPSQ* | -0.0007* | -33.67 | -0.0006* | -8.79 |
| *TENURE* | 0.0092* | 5.21 | 0.011* | 3.18 |
| *WKSWK* | 0.0021* | 8.08 | 0.004* | 6.86 |
| *WHITE* | 0.065* | 9.81 | 0.023 | 1.30 |
| *PRIVEM* | 0.186* | 15.79 | 0.252* | 9.49 |
| *GOVEM* | 0.145* | 11.21 | 0.224* | 7.24 |
| *SELFEM* | -0.060* | -4.46 | -0.028 | -0.92 |
| *US* | 0.010** | 1.95 | 0.057*** | 1.52 |
| *MARRIED* | 0.071* | 11.48 | -0.040* | -3.15 |
| *NORTH* | 0.117* | 3.36 | 0.295* | 3.17 |
| *SOUTH* | 0.164* | 4.72 | 0.330* | 3.53 |
| *MOUNT* | 0.144* | 4.03 | 0.305* | 3.19 |
| *PACIFIC* | 0.184* | 5.26 | 0.346* | 3.69 |
| $\lambda$ | -0.010* | -2.95 | -0.0085** | -1.78 |
| $R^2$ | 0.1359 | | 0.1027 | |
| Adj. $R^2$ | 0.1355 | | 0.1006 | |
| F | 382.086* | | 48.901* | |
| Sample Size | 38899 | | 6853 | |

Note: *indicates statistical significance at the 1 percent level.
**indicates statistical significance at the 5 percent level.
***indicates statistical significance at the 10 percent level.

As explained previously, the human capital theory suggests that an individual's wage rate will increase rapidly early in his or her life as human capital is accumulated through formal schooling and on-the-job training, it will reach its maximum near the point in mid-life when human capital is at its largest, and it will eventually decline as the effects of reduced human capital accumulation are outweighed by the effects of depreciation in human capital. Thus the human capital theory suggests that the earnings equations should be quadratic in experience, implying that they should include both *EXP* and *EXPSQ* variables, with a positive sign expected for the coefficient of variable *EXP* and a negative sign for the coefficient of variable *EXPSQ*.

It is seen from Table 3 that for both men and women the coefficients of *EXP* and *EXPSQ* are positive and negative, respectively, as expected, and highly significant. The level of experience for which ln monthly earnings (*W*) is maximized is 25 years for men and 30 years for women.

Becker (1964) distinguishes between two types of training: General training and specific training. General training is defined as the type of training that, once acquired, is equally useful in all other firms. That is, general training increases the productivity of a worker in any firm. Specific training, on the other hand, is the type of training that increases productivity only in the firm it is acquired. That is, it is usable only in the firm providing the training, and the value of the training is lost once the worker leaves the firm. It is seen from Table 3 that the coefficient of *TENURE* is positive and significant for both men and women. This would indicate that firm-specific training is important in the case of highly educated professional employees.

The significant positive coefficient on *MARRIED* variable for men is consistent with the finding that a married man with responsibilities for his family is more productive and earns more than a single man.[13] On the other hand, the significant negative coefficient on the same variable for women would reflect the fact that a married woman with children (or with a possibility of having children) earns less than a single woman even if she continues to work (possibly because she is thought to be less reliable and less open to carrier opportunities).

The coefficient of *US* is positive and significant for both groups. This implies that being a U.S. citizen is an advantage in the labor market. The coefficient of *WHITE* is also positive for both groups but significant only for men. This may be reflecting the effects of discrimination against nonwhite men in the market.

The coefficients of *PRIVEM* and *GOVEM* are highly significant and positive for both men and women. Besides, the magnitudes of these coefficients are relatively large. These indicate that private and government employees earn more compared to self-employed and other workers. Finally, the coefficients of the locational variables are positive and highly significant implying that locational effects are important in determining the wages of professional employees.

Table 4 presents the decomposition of male-female earnings differential according to the four decomposition methods explained in Section II: (1) The method in which the nondiscriminatory wage structure is the male wage structure (Equation (8)); (2) the method in which the nondiscriminatory wage structure is the female wage structure (Equation (9)); (3) the method proposed by Cotton (1988) in which the nondiscriminatory wage structure is the weighted average of the wage structures for males and females; and (4) the method proposed by Neumark (1988) in which the nondiscriminatory wage structure is calculated from an earnings function estimated over the pooled sample of males and females. The calculated (observed) log earnings differential equals 0.462 (i.e., the left-hand side of the decomposition equations is equal to 0.462). The decomposition of this magnitude into explained and discrimination components under the four decomposition methods is shown in Table 4.

**Table 4:** Decomposition of the Male-Female Earnings Differential

| **Nondiscriminatory Wage Structure:** | **The Percentage of the Earnings Differential Due to:** | |
|---|---|---|
| | **Endowments** | **Discrimination** |
| Male $(b_m)$ | 25.8 | 74.2 |
| Female $(b_f)$ | 20.3 | 79.7 |
| Cotton (Weighted) $(b^* = l_m b_m + l_f b_f)$ | 24.9 | 75.1 |
| Neumark (Pooled) $(b^* = (X'X)^{-1}(X'Y) = b)$ | 33.7 | 66.3 |

Using the male wage structure as the nondiscriminatory wage structure, it is estimated that 74.2 percent of the earnings differential between men and women is due to discrimination, while using the female wage structure as the nondiscriminatory wage structure leads to an estimate of 79.7 percent. When the nondiscriminatory wage structure is the weighted average of the wage structures for males and females, 75.1 percent of the differential is attributable to wage discrimination. As noted by Oaxaca and Ransom (1994: 13-14), because the Cotton method estimates the nondiscriminatory wage structure as a convex linear combination of the male and the female wage structures, it is constrained to yield estimates that lie inside the bounds derived from assuming that the male or the female wage structure is the competitive structure. Finally, when the nondiscriminatory wage structure is the OLS estimate obtained from the pooled sample of males and females (the Neumark method), the portion of the earnings differential attributable to discrimination drops to 66.3 percent.

## V. CONCLUSION

Using data from a 1986 National Science Foundation survey, this paper has examined the extent of gender discrimination in the case of scientists and engineers. The standard Oaxaca/Blinder decomposition method and the Cotton/Neumark extensions of this method have been employed to decompose the male-female earnings differential. In addition, the study has accounted for the sample selection bias by using Heckman's two-stage, two-equation procedure. The results of the study have shown that after adjusting for a variety of wage-generating characteristics, a substantial amount of the earnings differential remains that can be attributable to pay discrimination. Although the four decomposition methods employed have produced different results, the discrimination portion of the observed earnings differential has always been significant. Thus it can be concluded that even the wages of highly educated and highly paid professionals are unfairly conditioned by sex. This result is consistent with the findings of previous studies.

## REFERENCES

Arrow, K. (1973), "The Theory of Discrimination," in *Discrimination in Labor Markets*, O. Ashenfelter and A. Rees (eds.), Princeton: Princeton University Press, 3-33.

Becker, G. S. (1957) *The Economics of Discrimination*, Chicago: University of Chicago Press.

Becker, G. S. (1962), "Investment in Human Capital: A Theoretical Analysis," *Journal of Political Economy* (Supplement), 70, 9-49.

(1964), *Human Capital: A Theoretical and Empirical Analysis, With Special Reference to Education*, New York: Columbia University Press for NBER.

Ben-Porath, Y. (1967), "The Production of Human Capital and the Life Cycle of Earnings," *Journal of Political Economy*, 75, 352-365.

Blau, F.D. (1984), "Discrimination Against Women: Theory and Evidence," in *Labor Economics: Modern Views*, W. Darity, Jr. (ed.), Boston: Kluwer-Nijhoff Publishing, 53-89.

Blinder, A.S. (1973), "Wage Discrimination: Reduced Form and Structural Estimates," *Journal of Human Resources*, 8, 436-455.

Bloom, D.E. and M. R. Killingsworth (1982), "Pay Discrimination Research and Litigation: The Use of Regression," *Industrial Relations*, 21, 318-339.

Bowen, W. and T.A. Finegan (1969), *The Economics of Labor Force Participation*, Princeton: Princeton University Press.

Cain, G.G. (1986), "The Economic Analysis of Labor Market Discrimination: A Survey," in *Handbook of labor economics*, O. Ashenfelter and R. Layard (eds.), Amsterdam: Elsevier Science Publishers, B.V., 603-785.

Cotton, J. (1988), "On the Decomposition of Wage Differentials," *Review of Economics and Statistics*, 70, 236-243.

Dolton, P. and G. Makepeace (1987), "Interpreting Sample Selection Effects," *Economics Letters*, 24, 373-379.

Gordon, N.M., Morton, T.E. and I.C. Braden (1974), "Faculty Salaries: Is There Discrimination by Sex, Race and Discipline?" *American Economic Review*, 64, 419-427.

Greene, W.H. (1993), *Econometric Analysis*, New York: Macmillan Publishing Company.

Gunderson, M. (1989), "Male-Female Wage Differentials and Policy Responses," *Journal of Economic Literature*, 27, 46-72.

Hamermesh, D.S. and A. Rees (1993), *The Economics of Work and Pay*, New York: Harper Collins College Publishers.

Heckman, J.J. (1979), "Sample Selection Bias as a Specification Error," *Econometrica*, 47, 153-161.

Juhn, C., Murphy, K. and B. Pierce (1991), "Accounting for the Slowdown in Black-White Wage Convergence," in *Workers and Their Wages: Changing Patterns in the United States*, M. Kosters (ed.), Washington, D.C.: American Enterprise Institute Press, 107-143.

Korenman, S. and D. Neumark (1991), "Does Marriage Really Make Men More Productive?", *Journal of Human Resources*, 26, 282-307.

Loury, L.D. (1997), "The Gender Earnings Gap Among College-Educated Workers," *Industrial and Labor Relations Review*, 50, 580-593.

Madden, J.F. (1985), "The Persistence of Pay Differentials: The Economics of Sex Discrimination," in *Women and Work*, L. Larwood, A. H. Stromberg, and B. A. Gutek (eds.), Beverly Hills: Sage Publications, 76-114.

Malkiel, B.G. and J. A. Malkiel (1973), "Male-Female Pay Differentials in Professional Employment," *American Economic Review*, 63, 693-705.

Mincer, J. (1958), "Investment in Human Capital and Personal Income Distribution," *Journal of Political Economy*, 66, 197-201.

(1962), "On-the-Job Training: Costs, Returns, and Some Implications," *Journal of Political Economy*, 70, 50-79.

(1974), *Schooling, Experience, and Earnings*, New York: NBER.

Neumark, D. (1988), "Employer's Discriminatory Behavior and the Estimation of Wage Discrimination," *Journal of Human Resources*, 23, 279-295.

Oaxaca, R. L. (1973), "Male-Female Wage Differentials in Urban Labor Markets," *International Economic Review*, 14, 693-709.

Oaxaca, R. L. and M. R. Ransom (1994), "On Discrimination and the Decomposition of Wage Differentials," *Journal of Econometrics*, 61, 5-21.

Osterman, P. (1979), "Sex Discrimination in Professional Employment: A Case Study," *Industrial and Labor Relations Review*, 32, 451-464.

Schultz, T.W. (1960), "Capital Formation by Education," *Journal of Political Economy*, 68, 571-583.

Schultz, T.W. (1961), "Investment in Human Capital," *American Economic Review*, 51, 1-17.

Terrell, K. (1992), "Female-Male Earnings Differentials and Occupational Structure," *International Labour Review*, 131, 387-404.

Weinberger, C. (1998), "Race and Gender Wage Gaps in the Market for Recent College Graduates," *Industrial Relations*, 37, 67-84.

Wood, R.G., Corcoran, M. E. and P. N. Courant (1993), "Pay Differences Among the Highly Paid: The Male-Female Earnings Gap in Lawyer's Salaries," *Journal of Labor Economics*, 11, 417-441.

Willis, R.J. (1986), "Wage Determinants: A Survey and Reinterpretation of Human Capital Earnings Functions," in *Handbook of labor economics*. O. Ashenfelter and R. Layard (eds.), Amsterdam: Elsevier Science Publishers B.V., 525-602.

Zveglich, J., Rodgers, Y. and W. Rodgers (1997), "The Persistence of Gender Earnings Inequality in Taiwan, 1978-1992," *Industrial and Labor Relations Review*, 50, 594-609.

## NOTES

1. For studies reviewing the theoretical and empirical literature on sex discrimination, see, for example, Terrel (1992), Gunderson (1989), Cain (1986), Madden (1985), and Blau (1984).
2. Among these are Weinberger (1998), Loury (1997), Wood, Corcoran and Courant (1993), Osterman (1979), Gordon, Morton and Braden (1974), and Malkiel and Malkiel (1973).
3. The other reasons for analyzing restricted samples include a concern about noncompeting groups, and an interest in within-group as opposed to workplace-wide discrimination. These are discussed in Bloom and Killingsworth (1982).
4. The outline of the decomposition procedure given below is taken from Gunderson (1989: 50-51).
5. The proportions of males and females in the whole regression sample can be taken as the appropriate weights to the extent that the male and female numbers in the sample represent the total male and female numbers in the population.
6 In an alternative decomposition approach, Juhn, Murphy and Pierce (1991) use only the male wage structure (and standard deviation from the male regression). The most prominent argument for using male coefficients is that they more accurately reflect competitive returns to wage generating characteristics than do female coefficients. This argument is held by most of the studies using male coefficients. Although the decomposition of Juhn, Murphy and Pierce differs from Equation (8) in its interpretation, it can be shown that when evaluated at the means, the two decompositions are identical with respect to the sizes of explained and discrimination components of the earnings differential. For a brief exposition of this see Zveglich, Rodgers and Rodgers (1997: 601).
7. The inverse of Mill's ratio for the *i*th individual is given by $\lambda_i = \phi(Z_i)/[1-\Phi(Z_i)]$ where $\phi$ and $\Phi$ are, respectively, standard normal density and cumulative distribution functions; where $Z_i = -Y_i\gamma/\sigma$ and where $\gamma$ is the vector of coefficients in the probit equation predicting inclusion in the wage regression, $Y_i$ is the vector of exogenous

variables in the probit equation, and $\sigma$ is the standard deviation of the disturbance term in the probit equation. For details see Heckman (1979).

8. In fact this specification is a modified version of Mincer's (1974) human capital earnings function. In Mincer's function earnings are determined by education, work experience and the square of work experience (see Willis, 1986 for a survey). The human capital earnings function is based on the human capital theory developed primarily by Becker (1962, 1964), Schultz (1960, 1961) and Mincer (1958, 1962). This function is also compatible with more formal models of human capital accumulation such as the Ben-Porath (1967) model.

9. The simultaneous inclusion of *EDUC*, *EXP* and *TENURE* variables in the earnings equations might cause a problem of multicollinearity between these variables. The variance inflation factors calculated for these variables reveal that there is not a severe multicollinearity problem in our data.

10. In general, variables thought to influence an individual's (potential) wage rate and his or her reservation wage should be included in a labor force participation equation. Individuals with higher (potential) wage rates and lower reservation wages are more likely to be in the labor force. See, for example, Hamermesh and Rees (1993) and Bowen and Finegan (1969).

11. This does not cause any econometric problem even though $\lambda$ (to be calculated from probit estimates) will be included as an independent variable in the earnings equations (see, for example, Greene, 1993, Ch. 22).

12. For both the probit estimates for labor force participation (Table 2) and the OLS estimates of earnings equations (Table 3), we do not include and discuss the estimation results for the pooled sample of males and females in order to save space.

13. See Korenman and Neumark (1991).