Bounding Estimates of Wage Discrimination

by

J. G. Hirschberg
Department of Economics
University of Melbourne
Parkville, 3052

and

D. J. Slottje
Department of Economics
Southern Methodist University
Dallas, TX 75275-0496

7 August 2002

Abstract:

The Blinder Oaxaca decomposition method for defining discrimination from the wage equations of two groups has had a wide degree of application. However, the implication of this measure can very dramatically depending on the definition of the non-discriminatory wage chosen for comparison. This paper uses a form of extreme bounds analysis to define the limits on the measure of discrimination that can be obtained from these decompositions. A simple application is presented to demonstrate the use of the bootstrap to define the distributions of the discrimination measure.

Key words: Extreme Bounds Analysis, Discrimination, Bootstrap

JEL Codes: J7, C2

0. Introduction

A rich literature on the empirical analysis of labor market discrimination has followed from the contributions of Blinder (1973) and Oaxaca (1973). These researchers were among the first to explore this issue econometrically. It has been understood for some time that the dichotomy in the average wages of two groups (usually broken down by sex or race and here referred to as the advantaged and the disadvantaged) is due in part to differences in average levels of productivity (or skill) (their endowment) and is due in part to disparate treatment of the two groups once they enter the labor market (the discrimination). However, the decomposition of the average wage differences into these two different parts has been found to vary with the method used. In this paper we propose a method for defining the bounds on these measures. Although recent contributions to the literature have investigated entry into the labor market and selectivity bias as additional reasons for the observation of large wage differentials this paper concentrates on the variation within the traditional Blinder-Oaxaca decomposition which for gender differences has recently been shown to be the most important element in the decomposition of wage differentials (for example see Madden 2000).

This paper proceeds as follows. First, we review the decomposition and the methods that have been proposed. Second we define the method for bounding the non-discriminatory wage parameters. Then we show how the measures of discrimination can be bounded. In the fourth section we operationalize the use of the bounds by providing approximations to the asymptotic variances of the discrimination measures. In Section five the bootstrap methods are defined for the estimation of the densities of the bounds on the discrimination measures. Section six defines a simple application using data that is widely available.

1. Decomposition of Wage Differences

Becker (1971) defined a measure of discrimination as the difference between the observed wage ratio and the wage ratio that would prevail in the absence of discrimination. This discrimination coefficient can be expressed as

$$\delta = \frac{\left(\frac{\overline{W}_a}{\overline{W}_d}\right) - \left(\frac{MP_a}{MP_d}\right)}{\left(\frac{MP_a}{MP_d}\right)} \tag{1}$$

where $\overline{W}_a$ is the average advantaged worker's wage in the market and $\overline{W}_d$ is the average disadvantaged worker's wage in the market. It is straightforward to see that

$$\frac{MP_a}{MP_d} = \left(\frac{\overline{W}_a}{\overline{W}_d}\right) \tag{2}$$

in the absence of discrimination and (2) follows from the usual cost minimization problem. Oaxaca (1973) introduced the formulation given in (1). Following Oaxaca (1973), Cotton (1988) noted that (1) can be written in logarithmic form

$$\ln \overline{W}_a - \ln \overline{W}_d = \ln MP_a - \ln MP_d + \ln(\delta + 1) \tag{3}$$

where the first term on the right hand side (the difference in the logs of the marginal products) is due to differences in productivity of the two groups and the second term on the right hand side (ln(δ+1)) is due to discrimination. Oaxaca (1973) showed that separate linear models of the log wage specification can be estimated for disadvantaged or *d*'s $\left(\ln(\overline{W}_d) = \overline{X}'_d \hat{\beta}_d\right)$ and advantaged or *a*'s $\left(\ln(\overline{W}_a) = \overline{X}'_a \hat{\beta}_a\right)$. The estimates can then be combined in the following way since regression lines must pass through the variables' means:

$$\ln(\overline{W}_a) - \ln(\overline{W}_d) = \overline{X}'_a \hat{\beta}_a - \overline{X}'_d \hat{\beta}_d \tag{4}$$

The formulation given in (4) follows Neumark's (1988) notation where $\bar{X}_a$ and $\bar{X}_d$ are vectors containing the means of the variables which are presumed to impact productivity (and subsequently wages) and $\hat{\beta}_a$ and $\hat{\beta}_d$ are the estimated coefficients. Empirical work using (4) has been done using two decompositions. If $\Delta X' = \bar{X}'_a - \bar{X}'_d$ and $\Delta\hat{\beta} = \hat{\beta}_a - \hat{\beta}_d$, then (4) becomes either,

(5) $$\ln(\bar{W}_a) - \ln(\bar{W}_d) = \Delta\bar{X}'_a\hat{\beta}_a + \bar{X}'_d\Delta\hat{\beta}$$

or

(6) $$\ln(\bar{W}_a) - \ln(\bar{W}_d) = \Delta\,\bar{X}'\hat{\beta}_d + \bar{X}'_a\Delta\hat{\beta}$$

where (5) and (6) are found by adding $\left(\bar{X}'_d\hat{\beta}_a - \bar{X}'_d\hat{\beta}_a\right)$ to (5) and adding $\left(\bar{X}'_a\hat{\beta}_d - \bar{X}'_a\hat{\beta}_d\right)$ to (6). The Oaxaca model decomposes the first term on the right hand side of (5) into the portion of the mean log wage differential due to differences in average productivity and the second term is due to different wage structures. The β's are given this interpretation since they reflect the returns that individuals will get from their personal characteristics with respect to wages. Unfortunately, as Neumark (1988) (among others) has pointed out, considerable variation may exist in the estimate one gets of the wage differential due to discrimination if one uses (5) vis á vis (6). Neumark (1988) presents a nice exposition on where the discrepancy lies in using (5) rather than (6) or vice versa. If (5) is selected as the model to detect discrimination, it is assumed the advantaged worker's wage structure becomes the one that would exist in the absence of discrimination. In (6), the disadvantaged worker's wage structure would be the prevailing one. These cases are both straightforward to see since without discrimination (where the second term would disappear in (5)), we would attribute the mean wage difference to differences in characteristics weighted by the advantaged workers wage structure (β_a). Neumark (1988) made this point even clearer by generalizing Oaxaca's result to get a broader decomposition:

(7) $$\ln(\overline{W}_a) - \ln(\overline{W}_d) = \Delta\overline{X}'\beta^* + \left[\ \overline{X}_a'\left(\hat{\beta}_a - \beta^*\right) + \overline{X}_d'\left(\beta^* - \hat{\beta}_d\right)\right]$$

where β^* is assumed to represent the wage structure that would prevail in the absence of discrimination. Neumark (1988) shows that (5) or (6) can be generated as special cases of (7) and thus emphasizes the import of what one assumes about β^* in attempting to measure discrimination. Cotton (1988) performed a similar analysis and argued that β^* should be constructed as a weighted average of advantaged and disadvantaged worker's wages weighted by the ratio of the disadvantaged to the advantaged labor force representation. Neumark (1988) rightly notes that this is an ad hoc specification and proposes finding β^* based on a more theoretical foundation.

Specifically, Neumark (1988) assumes the employer derives utility from profits and from the discrimination-based composition of the labor force. The utility function is assumed to be homogenous of degree zero with respect to the labor input. This means that if the numbers of the two groups of workers are changed proportionately, utility is unchanged. Neumark interprets this to mean that employers only care about the relative proportions of the two types of workers. Neumark's model ultimately leads to,

(8) $$MP_j = \frac{W_{aj}N_{aj} + W_{dj}N_{dj}}{N_{aj} + N_{dj}}$$

(where N_a is the number of advantaged workers and N_d is the number of disadvantaged workers) or that the marginal product of the *j*th worker depends on the relative proportions of the various types of labor so that since $W_j = MP_j$ in the absence of discrimination, the non-discrimination wage can be found from (8). Neumark (1988) finds the estimator of the non-discrimination wage structure (β^*) by first running regressions on the two sub-samples to get fitted log wage values and then after combining the fitted values of the log wages, by then running a regression on the whole sample. Those coefficient estimates will then give an estimate of β^*. One difficulty with the implementation of Neumark's method is that the sample used in estimation may not refect the number of employees a particular employer has hired in each category. It is quite common to apply these methods to data based on a sampling procedure that is not influenced by the employer's actions. Neumark's (1988) weighting procedure is similar to one used by Oaxaca and Ransom's (O-R) (1988) which was used in the context of estimating union wage effects. Oaxaca and Ransom (1991) also proposed a weighting matrix which was specified by

$$\Omega_N = (X'X)^{-1}(X_a'X_a) \tag{9}$$

where X is the observation matrix for the pooled (both classes of workers) sample and X_a is the observation matrix for the advantaged sample. The interpretation of Ω_N as a weighting matrix is readily seen by noting that $X'X = X_a'X_a + X_d'X_d$, where X_d is the observation matrix for the disadvantaged sample.

O-R showed that

$$\beta^* = \Omega_N\hat{\beta}_a + (\mathbf{I} - \Omega_N)\hat{\beta}_d \tag{10}$$

where β^* is the ordinary least squares estimator from the pooled sample (containing both types of workers.) Thus, this weighting scheme was found by O-R to be the ordinary least squares estimator from the combined groups as the wage structure that would exist in the absence of discrimination. They noted that this estimate of the common wage structure is not in general a convex, linear combination of the separately estimated advantaged and disadvantaged workers' wage structures and they get a result similar to that of Neumark.

As O-R note, Cotton's (1988) weighting is equivalent to O-R's when $\left({}^{N_a}\!/_{N}\right)(X'X)$ = $(X_a'X_a)$, if the first and second sample moments are identical for all workers. And because the sample mean characteristics for the advantaged and disadvantaged workers are the same, all of the differences in wages are due to discrimination.

To summarize the literature on the establishment of a hypothetical ideal (with no advantage or disadvantage given) wage structure (β^*) we summarize the findings in Table 1 in which we have identified the various definitions of Ω as proposed in previous research.

We now propose a different method for determining the extent to which the definition of β^* matters on the resulting definition of discrimination.

2. Bounding β^*

Leamer's 1978 monograph proposes a method for the determination of the fragility of a regression result. This is done by subjecting regression models to an analysis that determines the extreme bounds (EB) of parameter estimates based on the assumption of a prior distribution for selected parameters. In the usual application this is interpreted as a means for the comparison of all possible regression model specifications in which various subsets of regressors are considered for omission from the regression. The most widely cited example of this form of analysis can be found in Leamer's 1983 paper entitled "Let's take the con out of econometrics". Subsequently a number of papers have appeared that have criticized the EB approach to model specification analysis most notably McAleer Pagan and Volker (1985) as focusing on a very narrow type of specification choices and for the tendency for these analysis to reject too many models to be of much use. However, a resurgence of applications and modifications of Leamer's EB analysis have appeared in Levine and Renelt (1992), Gawande (1995), and Temple (2000) among a number of others. In this paper we do not use the EB analysis per say in that we do not investigate the implications of regression specification changes. However, we use one of the fundamental results on which EB analysis is based which allows us to define a bound all the possible parameter estimates that may be used for the nondiscriminatory wage structure. Then we solve an optimization problem that allows us to define two nondiscriminatory wage structures. One that will maximize the measure of discrimination and the other that will minimize the measure of discrimination.

Chamberlain and Leamer (1976) (C-L) consider the case of a vector β^* that can be defined as a matrix weighted average of two vectors

$$\beta^* = \left(\mathbf{H}_a + \mathbf{H}_d\right)^{-1}\left(\mathbf{H}_a\,\hat{\beta}_a + \mathbf{H}_b\,\hat{\beta}_d\right) \tag{11}$$

where the weighting matrices $\mathbf{H}_a$ and $\mathbf{H}_d$ are positive definite symmetric. In the applications they consider these two sets of parameters are identified in terms of a Bayesian estimator where one group would be identified as the data and the other as the prior with the resulting ideal or non-discriminatory set of parameters as the posterior and the **H**'s are the corresponding precision matrixes (or inverse covariance matrixes). Algebraically there is no distinction between the prior and the data though in practice Bayesian methods are often applied where detailed data distributions are defined but priors are non-informative.

In the case of the decompositions defined by Ω_{O}, Ω_{R} and Ω_{C} as defined in Table 1, we can set $\mathbf{H}_a = \Omega$ and $\mathbf{H}_d = \mathbf{I} - \Omega$. In the case of the Neumark decomposition $\mathbf{H}_a = X_a' X_a$ and $\mathbf{H}_d = X_d' X_d$ and the resulting (posterior) mean vector of parameters is equivalent to the Bayesian interpretation of the OLS estimator when there is an addition of data. Thus X_a would be added to X_d to form a total sample from which the estimate would be obtained.

$$\beta^* = \Omega\,\hat{\beta}_a + \left(\mathrm{I} - \Omega\right)\,\hat{\beta}_d \tag{12}$$

Where the matrix Ω is a positive definite symmetric matrix. Consequently, wage decompositions provide an application of methods developed for the consideration of these linear Bayesian models.

From Theorem 2 C-L prove that the matrix weighted average (β^*) must lie within the ellipsoid defined by $(\beta^* - c)'\mathbf{H}(\beta^* - c) < \frac{1}{4}\,\Delta\hat{\beta}'\mathbf{H}\Delta\hat{\beta}$. Where $c = \left(\hat{\beta}_d + \hat{\beta}_a\right)/2$ the arithmetic average of the parameter vectors and **H** is a sample precision matrix unique up to a scalar multiple. This provides a constraint on the extreme values of β^* as:

(13) $$(\beta^* - c)'\mathbf{H}(\beta^* - c) = ¼\Delta\hat{\beta}'\mathbf{H}\Delta\hat{\beta}$$

Which implies that any possible value of β^* defined by the different values of Ω must be contained within or on the surface of this ellipsoid.

From the relationship in (7) we have:

$$\ln(\overline{W}_a) - \ln(\overline{W}_d) = E + D$$

where:

(14) $$D = \left[\, \overline{X'}_a\left(\hat{\beta}_a - \beta^*\right) + \overline{X'}_d\left(\beta^* - \hat{\beta}_d\right)\right]$$

D is the difference in the log wages that is attributable to the differential payment schedule that is often referred to as "discrimination". Where the term $\overline{X'}_a\left(\hat{\beta}_a - \beta^*\right)$ measures the over compensation paid to the advantaged group and $\overline{X'}_d\left(\beta^* - \hat{\beta}_d\right)$ measures the under compensation paid to the disadvantaged group.

(15) $$E = \Delta\overline{X}'\beta^*$$

E is the difference that is due to the differences in the worker's characteristics/human capital which is referred to as "endowment". We can solve for the value of β^* as the value that either maximizes or minimizes D. By implication, since $\Delta\ln(\overline{W})$ remains constant, minimizing D maximizes E and maximizing D is equivalent to minimizing E. Thus we solve the following optimization problem:

(16) $$\text{Max/Min } \left(E = \Delta\overline{X}'\beta^*\right), \text{ st } (\beta^* - c)'\mathbf{H}(\beta^* - c) = ¼\Delta\hat{\beta}'\mathbf{H}\Delta\hat{\beta}$$

Where we use the full sample cross products matrix $X'X$ as the sample precision matrix $\mathbf{H}$ or the appropriate inverse of the heteroscedastic consistent covariance matrix. The constrained optimization can then be defined by a Lagrangian of the form:

(17) $$L = \Delta\overline{X}'\beta^* - \lambda\left((\beta^* - c)'\mathbf{H}(\beta^* - c) - ¼\Delta\hat{\beta}'\mathbf{H}\Delta\hat{\beta}\right)$$

The first order derivatives of L with respect to β^* and λ are given as:

(18) $$\frac{\partial L}{\partial \beta^*} = \Delta\overline{X} - 2\lambda\mathbf{H}(\beta^* - c)$$

(19) $$\frac{\partial L}{\partial \lambda} = (\beta^* - c)'\mathbf{H}(\beta^* - c) - ¼\Delta\hat{\beta}'\mathbf{H}\Delta\hat{\beta}$$

We can solve (18) for the optimal value of β^* ($\hat{\beta}^*$) by setting this expression equal to zero and we get:

(20) $$\hat{\beta}^* = c + \hat{\rho}\,\mathbf{H}^{-1}\Delta\overline{X} + \text{ ,where } \hat{\rho} = \frac{1}{2\hat{\lambda}}$$

then substituting $c + \hat{\rho}\,\mathbf{H}^{-1}\Delta\overline{X}$ for $\hat{\beta}^*$ into (19) which is also set to equal to zero we can solve for $\hat{\rho}$ where we get two solution vectors

(21) $$\hat{\rho} = \pm\hat{\phi}, \text{ where } \hat{\phi} = ½\sqrt{\frac{\Delta\hat{\beta}'\mathbf{H}\Delta\hat{\beta}}{\Delta\overline{X}'\mathbf{H}^{-1}\Delta\overline{X}}}$$

Then two solutions for the optimal β^* are found to be:

(22) $$\hat{\beta}_i^* = c + \gamma_i\phi\mathbf{H}^{-1}\Delta\overline{X}$$

where $\gamma_1 = 1$ and $\gamma_2 = -1$.

The second order conditions can be established by evaluating the matrix of second derivatives evaluated at each solution as:

(23) $$\frac{\partial^2 L\left(\beta_i^* \mid \hat{\lambda}\right)}{\partial\left(\beta^* \mid \lambda\right)^2} = -\gamma_i\left(\begin{array}{c|c} \hat{\phi}^{-1}\mathbf{H} & 2\hat{\phi}\Delta\overline{X} \\ \hline 2\hat{\phi}\Delta\overline{X}' & \mathbf{0} \end{array}\right)$$

Because the precision matrix $(\mathbf{H})$ is a positive definite matrix and $\hat{\phi} > 0$, β^*_1 will be the maximum of E and the minimum of D and β^*_2 will be the minimum value of E and the maximum of D and we can determine the bounds on the possible values of the measure of discrimination. Note that when $\beta_d = \beta_a$ then $\beta^* = \beta_d = \beta_a$.

3. Bounds on the measure of discrimination (*D*).

The extreme values of β_i^* can now be used to define the extreme values of the discrimination measure (*D*) which we will denote as D_i^*. From the definitions above we have that $\hat{D}_i^* = \Delta\ln(\overline{W}) - \Delta\overline{X}'\hat{\beta}_i^*$ or by substitution this can be shown to be:

$$\hat{D}_i^* = \Delta\ln(\overline{W}) - \Delta\overline{X}'\hat{\beta}_i^* \tag{24}$$

Thus

$$\hat{D}_i^* = \Delta\ln(\overline{W}) - \Delta\overline{X}'c - \gamma_i \tfrac{1}{2}\sqrt{\Delta\hat{\beta}'\mathbf{H}\Delta\hat{\beta}}\sqrt{\Delta\overline{X}'\mathbf{H}^{-1}\Delta\overline{X}} \tag{25}$$

recall that $\gamma_1 = 1$ and $\gamma_2 = -1$. Thus the difference between the limiting values of the discrimination measure is given by

$$\hat{D}_2^* - \hat{D}_1^* = \sqrt{\Delta\hat{\beta}'\mathbf{H}\Delta\hat{\beta}}\sqrt{\Delta\overline{X}'\mathbf{H}^{-1}\Delta\overline{X}} \tag{26}$$

which is a weighted function of differences in the vector of parameters ($\Delta\hat{\beta}$) and ($\Delta\overline{X}$). Thus the greater the difference in the parameters or the greater the difference in the discrimination measures the larger the span of values one might obtain from any discrimination measure employed.

The measure *D* can also be shown to be directly related to the measure of discrimination defined in (1) as δ. From the relationship in (7) and (14) and (15) we have:

$$\ln\left(\frac{\overline{W}_a}{\overline{W}_d}\right) = E + D \tag{27}$$

If we are interested in removing the influence of the differences in endowments, or equivalently making the assumption that $MP_a = MP_d$ we can concentrate on the value of *D*.

$$\ln\left(\frac{\overline{W}_a}{\overline{W}_d}\right) = D \tag{28}$$

or equivalently:

(29) $$\left(\frac{\overline{W}_a}{\overline{W}_d}\right) = \exp(D)$$

as the ratio of the average wage for the advantaged group to the disadvantaged group. And we define:

(30) $$\begin{aligned}\overline{W_a} &= \exp(D)\overline{W_d} \\ \overline{W_a} &= (1+\delta)\overline{W_d}\end{aligned}$$

by equation (1). Thus we have that:

(31) $$\delta = \exp(D) - 1$$

Or that δ is a monotonic function of *D* and the maximization of *D* will coincide with the maximum of δ and the minimization of *D* is also the minimum value of δ. Note that when $|D| < .3$ the approximation that $\delta \approx D$ can be used.

We can define the estimate of δ using any particular definition of $\hat{\beta}^*$ as:

(32) $$\hat{\delta}_i = \exp\left[\hat{D}_i^*\right] - 1$$

In order to use the estimated values of *D* and β^* to make inferences we need to be able to make probability statements concerning their estimates. A first step in making these inferences is the derivation of an estimate for their variances.

4. The asymptotic variance of $\hat{D}$ and $\hat{\beta}^*$

In a companion paper to their 1994 paper Oaxaca and Ransom (1998) present the methodology for the computation of the variances used in their earlier paper. The technique they employ is an application of the widely used "delta method" in which a first order Taylor series expansion is used to linearize *D*. In this section we also apply the delta method but we consider not only the estimated parameters but in a difference from Oaxaca and Ransom we also assume that the means of the characteristics of each group are stochastic as well. Thus *D* is defined in terms of four random vectors ($\hat{\beta}_a, \hat{\beta}_d, \overline{X}_a,$ and $\overline{X}_d$) for which we can define estimates of their covariances. By stacking these four vectors we define a vector of length $4k$ given as $\boldsymbol{\theta}$ which is defined as:

(33)
$$\hat{\boldsymbol{\theta}}' = \left[\ \hat{\beta}'_a \ \middle|\ \hat{\beta}'_d \ \middle|\ \bar{X}'_a \ \middle|\ \bar{X}'_d\ \right]_{1\times 4k}$$

Where the covariance of $\hat{\boldsymbol{\theta}}$ is defined as $\boldsymbol{\Psi}$ and we can define this covariance as:

(34)
$$\boldsymbol{\Psi} = \left[\begin{array}{c|c|c|c} \Phi_a & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \hline \mathbf{0} & \Phi_d & \mathbf{0} & \mathbf{0} \\ \hline \mathbf{0} & \mathbf{0} & \Sigma_a & \mathbf{0} \\ \hline \mathbf{0} & \mathbf{0} & \mathbf{0} & \Sigma_d \end{array}\right]_{4k\times 4k}$$

The estimates of Σ_i are the covariances of the means of the attributes for each group and the $\Phi_i = \operatorname{cov}\left(\hat{\beta}_i\right)$ is the appropriate estimator of the parameter covariance matrix which may need to be corrected to account for heteroskedasticity, a commonly encountered problem in the estimation of wage equations, or may be the product of a maximum likelihood estimation in the case that the earnings data are not provided in continuous records.

In order to estimate the variance of the measure of discrimination we use the delta method which results in:

(35)
$$\widehat{\operatorname{var}}\left(\hat{D}\right) = \left[\frac{\partial D(\hat{\theta})}{\partial \theta}\right]' \hat{\boldsymbol{\Psi}} \left[\frac{\partial D(\hat{\theta})}{\partial \theta}\right]$$

Consequently this estimate requires the definition of the gradient of D with respect to the parameters in θ. For the previously defined set of discrimination measures defined in Section 1 of this paper, as determined by the weighting matrix Ω (as summarized in Table 1), we find the following estimate of the variance:

$$
\begin{aligned}
\widehat{\text{var}}\left(\hat{D}\right) &= \left(\hat{\beta}_a - \beta^*\right)' \hat{\Sigma}_a \left(\hat{\beta}_a - \beta^*\right) \\
&+ \left(\hat{\beta}_d - \beta^*\right)' \hat{\Sigma}_d \left(\hat{\beta}_d - \beta^*\right) \\
&+ \left(\bar{X}_a - \Omega' \Delta\bar{X}\right)' \hat{\Phi}_a \left(\bar{X}_a - \Omega' \Delta\bar{X}\right) \\
&+ \left(\bar{X}_d - \left(\mathbf{I} - \Omega\right)' \Delta\bar{X}\right)' \hat{\Phi}_d \left(\bar{X}_d - \left(\mathbf{I} - \Omega\right)' \Delta\bar{X}\right)
\end{aligned} \tag{36}
$$

In the case of the extreme values of D that we have derived in Section 2 we do not define a unique value for the weighting matrix Ω. Thus β^* is not a linear function of the parameter estimates for each case ($\hat{\beta}_a$ and $\hat{\beta}_d$) consequently we need to derive a different expression for the approximate variance based on the equation (25) given as:

$$
\begin{aligned}
\widehat{\text{var}}(D_i^*) &= \left[\hat{\beta}_a + c - \gamma_i \hat{\rho} \mathbf{H}^{-1} \Delta\bar{X}\right]' \hat{\Sigma}_a \left[\hat{\beta}_a + c - \gamma_i \hat{\rho} \mathbf{H}^{-1} \Delta\bar{X}\right] \\
&+ \left[-\hat{\beta}_d - c + \gamma_i \hat{\rho} \mathbf{H}^{-1} \Delta\bar{X}\right]' \hat{\Sigma}_a \left[-\hat{\beta}_a + c + \gamma_i \hat{\rho} \mathbf{H}^{-1} \Delta\bar{X}\right] \\
&+ \left[\bar{X}_a - \tfrac{1}{2}\Delta\bar{X} - \gamma_i \tfrac{1}{4}\hat{\rho}^{-1} \mathbf{H} \Delta\hat{\beta}\right]' \hat{\Omega}_a \left[\bar{X}_a - \tfrac{1}{2}\Delta\bar{X} - \gamma_i \tfrac{1}{4}\hat{\rho}^{-1} \mathbf{H} \Delta\hat{\beta}\right] \\
&+ \left[-\bar{X}_d - \tfrac{1}{2}\Delta\bar{X} + \gamma_i \tfrac{1}{4}\hat{\rho}^{-1} \mathbf{H} \Delta\hat{\beta}\right]' \hat{\Omega}_a \left[-\bar{X}_d - \tfrac{1}{2}\Delta\bar{X} + \gamma_i \tfrac{1}{4}\hat{\rho}^{-1} \mathbf{H} \Delta\hat{\beta}\right]
\end{aligned} \tag{37}
$$

again where $\gamma_1 = 1$ and $\gamma_2 = -1$.

In addition, we can define the approximate covariance of both of the extreme value parameters ($\hat{\beta}_1^*$ and $\hat{\beta}_2^*$), as defined in equation (22) as:

$$
\begin{aligned}
\widehat{\text{cov}}(\hat{\beta}_i^*) &= \hat{\rho}^2 \mathbf{Q}' \mathbf{H}^{-1} \left[\hat{\Sigma}_a + \hat{\Sigma}_d\right] \mathbf{H}^{-1} \mathbf{Q} \\
&+ \tfrac{1}{4} \left\{ \left[\mathbf{I} + \gamma_i \mathbf{G}\right]' \left[\hat{\Phi}_a\right] \left[\mathbf{I} + \gamma_i \mathbf{G}\right] + \left[\mathbf{I} - \gamma_i \mathbf{G}\right]' \left[\hat{\Phi}_d\right] \left[\mathbf{I} - \gamma_i \mathbf{G}\right] \right\}
\end{aligned} \tag{38}
$$

where $\mathbf{Q} = \mathbf{I} - \Delta\bar{X} \Delta\bar{X}' \mathbf{H}^{-1}$, $\mathbf{G} = \hat{\pi} \mathbf{H} \Delta\hat{\beta} \Delta\bar{X}' \mathbf{H}^{-1}$, $\hat{\rho}^2 = \tfrac{1}{4}\left(\Delta\hat{\beta}' \mathbf{H} \Delta\hat{\beta}\right)\left(\Delta\bar{X}' \mathbf{H}^{-1} \Delta\bar{X}\right)^{-1}$, and

$\hat{\pi} = \left(\Delta\hat{\beta}' \mathbf{H} \Delta\hat{\beta}\right)^{1/2} \left(\Delta\bar{X}' \mathbf{H}^{-1} \Delta\bar{X}\right)^{1/2}$.

5. Bootstrapping standard errors and confidence intervals for *D*

An alternative to constructing the Wald tests using the approximate variances defined in (37) and (38) is to employ Efron's (1982) bootstrap to construct alternative standard error estimates and confidence intervals that are not based on any particular distribution. The bootstrap has been applied in the computation of discrimination measures most notably by Silber and Weber (1999) where they compare the values for the discrimination measures defined in Table 1 for the differences between "Easterners" and "Westerners" in the Israeli labor market.

The bootstrap involves the recomputation of multiple values of the coefficients of interest ($\hat{D}_i^*$ and $\hat{\beta}_i^*$) by drawing with replacement from the data used. Since Efron's original contribution a number of enhancements have been proposed to the bootstrap methodology. In difference to Silber and Weber who employ the naive percentile approach on the measure of discrimination, we follow Horowitz's (2001) advice to base the bootstrap only on a pivot statistic. We use a conditional bootstrap for the regression coefficients as proposed in Freedman and Peters (1984) in which the model is assumed but the regression errors are sampled with replacement. The confidence intervals are constructed using a bootstrap-*t* technique as described in Efron and Tibshirani (1993) which is equivalent to using the asymptotic *t*-statistic as our pivot. The sampling with replacement is conducted using a second-order balanced resample method proposed by Davison, Hinkley and Schechtman (1986). This means that the average characteristics of each group ($\overline{X}_a$ and $\overline{X}_d$) are both resampled using the same sample as the residuals used to recompute the parameter estimates ($\hat{\beta}_a$ and $\hat{\beta}_d$). In addition, these samples are drawn in such a way to insure that the frequency of choosing each observation is equal.

In the case of the measures of discrimination D we use the t-ratio of the estimate to the estimated standard error as defined in (36) and (37) to form the appropriate pivot statistic. A statistic defined as a t-statistic is computed for each bootstrap simulation which is defined as:

$$t_b = \left(\hat{D}_b - \hat{D}\right) \Big/ \sqrt{\widehat{\text{var}}(\hat{D}_b)} \tag{39}$$

where the $\hat{D}_b$ denotes the estimated discrimination measure for bootstrap simulation (b) and $\hat{D}$ is the point estimate based on the data. These statistics are then rescaled to generate a bootstrap-t value of the discrimination measure designated as $\tilde{D}_b$ which is defined as:

$$\tilde{D}_b = \left(t_b \sqrt{\widehat{\text{var}}(\hat{D})}\right) + \hat{D} \tag{40}$$

6. A Simple Example

The differences in average wages for men and women in the US has been well documented. A number of papers have shown how this differential has changed over time in the US indicating that the differential has been decreasing over time (see Polachek and Robust 2001). The example we use here computes the various measures of discrimination as we have defined in the context of males as the advantaged group and women as the disadvantaged group. We use a small random subset of the 1985 Current Population Survey (245 women and 289 men) from Berndt(1991) (CPS85 from the data for chapter 5). Two regressions are estimated by gender, with the log of income as the dependent variable and the years of education and potential experience (as approximated by the number of years since left school) as the independent variables. The mean and standard deviation of the data are listed in Table 2. The regression parameter estimates are listed in Table 3. From these regressions we find that men are compensated at almost double the rate for their potential experience than women (.0163 versus .0089) although education seems to be better accounted for in women.

In Table 4 we list the various measures of discrimination (in terms of the log of the income). The differences of the means of the log of wages which includes both the endowment differences and the difference attributable to discrimination is found to be .2313. From the rest of the rows in Table 4 we find that all of the point estimates of the measures of discrimination are larger than this value which would indicate that the endowment has a negative effect on the wage difference. This table includes the point estimate in the 3 column and the approximate standard error in column 4. In addition, we have included the bootstrapped values of the mean, standard error, and the 95% confidence bounds. Note that for the traditional measures of discrimination the D_d to D_n measures the point estimate and the mean of the bootstrap estimates are very close indicating little bias. Also the asymptotic standard error estimates are almost exactly equal to the bootstrap values. In the bootstraps performed here we used 10,000 replications once we determined that more replications did not effect the results obtained to any significant degree.

Table 5 lists the extreme bounds for the parameter estimates (β_i^*) along with the asymptotic standard error estimates. We see that the non-discriminatory wage parameters that maximize the discrimination are those that result in parameters for potential experience that are small and for which we could not reject the hypothesis that they are equal to zero. And for the minimum set of non-discriminatory parameters are those that have the greatest parameter for the influence of potential experience and for education as well. In the last two rows of Table 4 we list the discrimination measures based on the bounds of the non-discriminatory wage parameters (β_i^*). Note that $D_1^* < [D_d \to D_a] < D_2^*$, the upper and lower bound estimates act as the limits on the estimates of the all the alternative discrimination measures. In this example, the extreme measures the asymptotic and bootstrap values differ more than for the other measures. The average of the bootstrapped values indicates that the point estimate of D_1^* (based on the minimum for the discrimination measure) may be positively biased and D_2^* (based on the maximum for the discrimination measure) may be negatively biased, though in neither case is the estimated bias more than 5%. From the bootstrapped confidence intervals we find that the 2.5% lower bound for the minimum value of the discrimination measure is .1545 and the 97.5% upper bound for the maximum of the discrimination measure is .3700. Thus we can bound the estimate of the discrimination measure although these probability statements ignore the probability of choice between the two extremes and any variation that may be due to alternative model specifications.

An equivalent method for demonstrating the probability bounds for the discrimination measure is by examining the density of the two extreme measures. Figure 1 displays two kernel density estimates as determined by the 10,000 studentized bootstrap values for each measure. Note that the density estimate for the lower bound appears to be estimated with greater precision than the upper bound as was the case for the bootstrapped variance estimate as borne out by the bootstrap estimate of the standard deviation for D_1^* as opposed to the standard deviation estimate for D_2^*. However it is apparent from this figure that the examination of the minimum discrimination measure results in an unambiguous conclusion that discrimination is non-zero in this case. In other words we could reject the hypothesis that discrimination was zero with a very low probability of making an error. Thus by using the minimum measure of discrimination and the lowest bound we still find that discrimination is positive.

A caveat for this application is in order. The model specification may create a larger degree of measured discrimination due to the lack of more detail as to education type, occupation, characteristics of the employer, family circumstances, and the proxy for experience. In particular, the use of potential experience alone for both men and women is probably responsible for increasing the measured discrimination due to the inadequacy of this variable to account for the differential in accumulated human capital that has been shown to explain such a large proportion of the gender wage gap (see Polachek 1995). Filer (1993) demonstrates empirically that this is an inappropriate proxy for a comparable experience measure for both men and women by demonstrating how other proxies change the gender differentials in coefficients. Specifically potential experience does not account for potential gaps in experience which are more prevalent for married women and women with children than for men. By measuring less actual experience for women than for men it is expected that the parameter in a wage equation would be less as well.

7. Conclusions

It is well known that the various wage differential decompositions traditionally done in analyzing discrimination rely heavily on the assumption regarding the non discrimination wage structure β^* (see equation (7)). Several authors have attempted to motivate the specification of this "no discrimination" wage structure based on the objective function of the employer in practicing discriminatory behaviour. The purpose of this paper has been to show that the wage structure that would prevail in the absence of discrimination can in fact be bounded when we assume that the information to establish this wage structure is a weighted average of the wage structure for the advantaged and the disadvantaged groups. Based on a theorem from Chamberlain and Leamer (1976) we showed in this paper that the non-discrimination wage parameters (β^*) must lie within an ellipsoid defined by the data and the regression results for each group. By using this method we are able to select the β^* which will maximize (minimize) the level of the discrimination in the labor market.

In addition to deriving the formulas for the estimated parameters for the non-discrimination wage structure that minimizes the level of discrimination we also specify the approximate standard errors. The point estimate and the approximate standard errors can be used to define a pivot statistic which can be used to bootstrap the discrimination measures. Thus it is possible to construct an estimate of the density of the discrimination measures which can then be used to make probability statements concerning the presence of discrimination. In the example used here we found that the measure of discrimination that was constructed was unambiguously positive as defined by the distribution of both the minimum discrimination measure.

REFERENCES

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

Chamberlain, G. and E. Leamer, "Matrix Weighted Averages and Posterior Bounds," *Journal Royal Statistical Society*, Series B, 38, (1976), 73-84.

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

Efron, B., *The Jackknife, the Bootstrap and Other Resampling Plans*, Society for Industrial and Mathematics, (1982).

Filer, Randall K., "The Usefulness of Predicted Values for Prior Work Experience in Analyzing Labor Market Outcomes for Women", *The Journal of Human Resources,* 28 (3), (1993), 519-537.

Freedman, D.A. and S.C. Peters, "Bootstrapping a Regression Equation: Some Empirical Results", *Journal of the American Statistical Association*, 79, (1984), 97-106.

Gawande, K., "Are U.S. Nontariff Barriers Retaliatory? An Application of Extreme Bounds Analysis in the Tobit Model", *The Review of Economics and Statistics*, 77, (1995), 677-688.

Leamer, E. E., "Let's Take the Con Out of Econometrics", *The American Economic Review*, 73, (1983), 31-43.

Levine, R. and D. Renelt, "A Sensitivity Analysis of Cross-Country Growth Regressions", *The American Economic Review*, 82, (1992), 942-963.

McAleer, M. A. R. Pagan and P. A. Volker, "What Will Take the Con Out of Econometrics?", *The American Economic Review*, 75, (1985), 293-307.

Madden, David, "Towards a broader explanation of male-female wage differences", *Applied Economics Letters*, 7, (2000), 765-770.

Madden, J. F., "Discrimination—A Manifestation of Male Market Power?" in Cynthia B. Lloyd (ed.), *Sex, Discrimination, and the Division of Labor* (New York: Columbia University Press, 1975).

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

Oaxaca, R., "Male-Female Wage Differentials in Urban Labor Markets," *International Economic Review*, 9, (Oct. 1973), 693-709.

Oaxaca, R. and M. Ransom, "Searching for the Effect of Unionism on the Wages of Union and Nonunion Workers," *Journal of Labor Research*, 9, (Spring 1988), 139-148.

______, "On Discrimination and the Decomposition of Wage Differentials," *Journal of Econometrics*, 61, (March 1994), 5-21.

______, "Calculation of Approximate Variances for Wage Decomposition Differentials", *Journal of Economic and Social Measurement*, 24, (1998), 55-61.

Polachek, Solomon W., "Human Capital and the Gender Earnings Gap: A Response to Feminist Critiques", in *Out of the Margin: Feminist Perspectives on Economics,* edited by Edith Kuiper and Jolande Sap, Routledge, 1995, 61-79.

Polachek, Solomon W. and John Robust, "Trends in the Male-Female Wage Gap: The 1980s Compared with the 1970s, *Southern Economic Journal*, 67(4), (2001), 869-888.

Reimers, C., "Labor Market Discrimination Against Hispanic and Black Men," *The Review of Economics and Statistics*, 65, (Nov. 1983), 570-579.

Silber, Jacques and Michal Weber, "Labour market discrimination: are there significant differences between the various decompositions?", *Applied Economics*, 31, (1999), 359-365.

Temple, J., "Growth Regressions and What the Textbooks Don't Tell You", *Bulletin of Economic Research*, 52, (2000), 181-205.

Table 1 The proposed values of the weighting matrix Ω.

Weighting Matrix	Author
$\Omega_O = \mathbf{I}$, or 0	Oaxaca (1973)
$\Omega_R = \frac{1}{2}\mathbf{I}$	Reimers (1983)
$\Omega_C = (N_a/N)\,\mathbf{I}$	Cotton (1988)
$\Omega_N = (X_a{'}X_a + X_d{'}X_d)^{-1}\,(X_a{'}X_a)$	Neumark (1988)

Table 2 The characteristics of the simple example.

Gender	Variable	Mean	SD
Men (289 obs)	natural logarithm of average hourly earnings	2.165	0.534
	potential years of experience (AGE-ED-6)	16.965	12.135
	years of education	13.014	2.768
Women (245 obs)	natural logarithm of average hourly earnings	1.934	0.492
	potential years of experience (AGE-ED-6)	18.833	12.613
	years of education	13.024	2.429

Table 3 Result of simple model regression

Gender	Variable	$\hat{\beta}$	SE	t-statistic
Men (R^2=.232, $\hat{\sigma}$ = .469)	(Constant)	0.7128	0.1614	4.4168
	potential years of experience (AGE-ED-6)	0.0163	0.0024	6.6904
	years of education	0.0903	0.0107	8.4298
Women (R^2=.262, $\hat{\sigma}$ = .423)	(Constant)	0.3110	0.1771	1.7564
	potential years of experience (AGE-ED-6)	0.0089	0.0023	3.8796
	years of education	0.1117	0.0119	9.3859

Table 4. Measures of discrimination with bootstrapped statistics based on simple model.

Variable	*Reference Parameters*	*Est*	*Asymptotic Std Dev,*	*Bootstrapped values*			
				Mean	**Std Dev**	**2.5%**	**97.5%**
$\Delta \ln(\bar{Y})$		.2313	.0446	.2313	.0452	.1456	.3182
D_d	$\hat{\beta}_d$	.2491	.0396	.2491	.0399	.1737	.3257
D_r	$½\left(\hat{\beta}_a+\hat{\beta}_d\right)$	.2559	.0391	.2559	.0394	.1812	.3321
D_a	$\hat{\beta}_d$	.2627	.0397	.2627	.0401	.1866	.3402
D_c	$\left(n_a\hat{\beta}_a+n_d\hat{\beta}_d\right)/\left(n_a+n_d\right)$	.2565	.0392	.2565	.0394	.1816	.3327
D_n	$\hat{\beta}$	.2543	.0391	.2543	.0392	.1800	.3302
D_1^*	β_1^*	.2327	.0549	.2287	.0437	.1545	.3025
D_2^*	β_2^*	.2790	.0473	.2831	.0462	.2005	.3700

Table 5 Extreme Bounds comparison parameter estimates ($\hat{\beta}_i^*$)

Bound	**Variable**	$\hat{\beta}$	**SE (asy)**	**t-statistic**
Min of D ($\hat{\beta}_1^*$)	(Constant)	0.0867	0.3950	0.2195
	potential years of experience (AGE-ED-6)	0.0229	0.0044	5.2631
	years of education	0.1196	0.0284	4.2095
Max of D ($\hat{\beta}_2^*$)	(Constant)	0.9367	0.3970	2.3596
	potential years of experience (AGE-ED-6)	0.0023	0.0042	0.5472
	years of education	0.0825	0.0286	2.8805

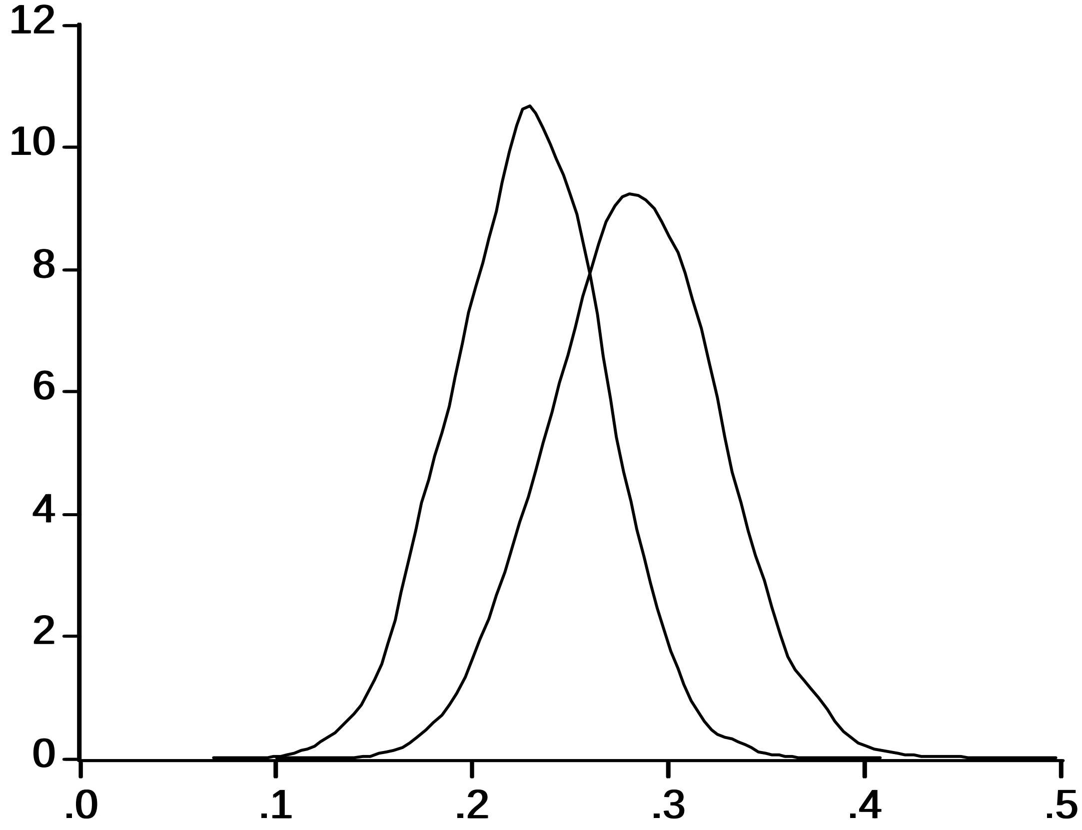

Figure 1. A comparison of the estimated densities of the t-bootstrapped values of D_1^* and D_2^*