DOCUMENT RESUME

ED 246 158 UD 023 681

AUTHOR	Abowd, John M.; Killingsworth, Mark R.
TITLE	Employment, Wages, and Earnings of Hispanics in the Federal and Non-Federal Sectors: Methodological Issues and Their Empirical Consequences.
SPONS AGENCY	Employment and Training Administration (DOL), Washington, D.C. Office of Research and Development.
PUB DATE	Sep 82
GRANT	21-36-78-61
NOTE	84p.; Also contained in UD 023 679.
PUB TYPE	Reports - Research/Technical (143)
EDRS PRICE	MF01/PC04 Plus Postage.
DESCRIPTORS	Blacks; Comparative Analysis; *Equal Opportunities (Jobs); *Ethnic Discrimination; Federal Government; Females; *Government Employees; *Hispanic Americans; Males; Racial Differences; Research Methodology; *Salary Wage Differentials; Sex Discrimination; *Statistical Analysis; Whites
IDENTIFIERS	*Private Sector; Survey of Income and Education

ABSTRACT

This paper has two purposes: (1) to examine whether Puerto Ricans, non-Puerto Rican Hispanics, and Blacks suffer substantial wage discrimination relative to comparable Whites; and (2) to examine the extent to which employers in the Federal and non-Federal sectors discriminate by race or ethnicity in making wage offers. After a discussion of economic theory underlying the statistical models employed, the 1976 Survey of Income and Education are subjected to direct, reverse, and structural regression analysis. In addition, data from the Federal Government Central Personnel Data File are subjected to direct and reverse regression analysis. Overall, the results of the three analysis techniques support the following findings: (1) minority women do not generally suffer substantial wage discrimination relative to comparable Whites; (2) minority men may suffer discrimination in terms of both wage offers and actual average wages, and estimates of the magnitude of both kinds of discrimination may be subject to serious measurement error bias; and (3) wage discrimination against minority males (particularly Blacks) is greater in the Federal than in the non-Federal sector, while earnings discrimination against minority males (particularly Blacks) is smaller in the Federal than in the non-Federal sector. (CMG)

Employment, Wages, and Earnings of Hispanics in the Federal and Non-Federal Sectors: Methodological Issues and Their Empirical Consequences

John M. Abowd
Graduate School of Business
University of Chicago
Economics Research Center/NORC

Mark R. Killingsworth
Department of Economics
Rutgers--The State University
Economics Research Center/NORC

This paper is based on a report prepared for the Employment and Training Administration, U.S. Department of Labor, under Research and Development Grant No. 21-36-78-61. Since grantees conducting research and development projects under government sponsorship are encouraged to express their own judgment freely, this paper does not necessarily represent the official opinion or policy of the Department of Labor. The authors are solely responsible for the contents of this paper and the associated report. We wish to acknowledge the assistance of Anthony Abowd, who performed most of the calculations for this paper and for the detailed report in conjunction with his Ph.D. thesis, in progress at the University of Chicago. Research assistance was provided by Paul McCudden and Leslie Brown of the University of Chicago.

Employment, Wages, and Earnings of Hispanics in the Federal and Non-Federal Sectors: Methodological Issues and Their Empirical Consequences

A major reason for studying employment and earnings differences by race and ethnicity is to determine what such differences imply both about potential employer discrimination and other sources of economic disadvantage resulting from race or ethnic origin. Much domestic policy is concerned with such questions, and information about the extent to which low economic status is related to employer discrimination or to other factors may have important implications for the allocation of resources to different domestic social programs such as antidiscrimination efforts, manpower training, and education programs.[1]

The results of statistical analyses of black/white and male/female wage and earnings differentials generally reveal that (1) on average, black and female wages and earnings are substantially below white male wages and earnings, and (2) even after adjustment for productivity-related factors such as schooling and labor force experience, the adjusted average level of black and female wages and earnings remains below the adjusted average level of white male wages and earnings. The difference between the adjusted average earnings or wages of blacks and of women and the adjusted average earnings or wages of white men is often called "labor market discrimination" to distinguish it from the differences in average earnings and wages that result from different levels of the productivity variables whose influence has been removed in the adjustment.

A major stylized fact that summarizes most of the empirical evidence on wage and earnings differentials is that both the black/white and the male/female adjusted differentials remain statistically and economically important regardless of the economic model or the statistical technique used to analyze the data. Specifically, black/white and male/female "labor market discrimination" have not been fully explained by either structural economic theories or statistical adjustments designed to eliminate a plethora of potential biases. In this paper we show that this stylized finding does not apply to Hispanic/Anglo wage and earnings differentials. Rather, on the whole, Hispanic/Anglo wage and earnings differences can generally be explained by human capital differences, self-selection biases, and statistical biases arising from imperfect measurement of the human capital differences. In particular, most of the difference between Hispanics and white non-Hispanics arises from human capital differences. A smaller but still important part of the difference arises from statistical biases due to measurement problems. Correcting for self-selection bias gives essentially the same results as ordinary regression analysis.

It is not possible to discuss literally all analytical and empirical questions about the sources of labor market differences in a single paper. Accordingly, we have limited the scope of our analyses in order to devote proper attention to (and to extend the range of analyses of) a number of specific issues. One issue to which we devote special attention is employer wage discrimination; another is the extent to which employers in the federal and non-federal sectors discriminate by race or ethnicity in making wage offers.[2]

Before proceeding, we define a number of concepts that figure prominently in what follows.

By "federal" and "non-federal" employment we mean, respectively, employment in the federal government and employment elsewhere in the economy.

By "ethnicity" we mean Hispanic or non-Hispanic ethnic origin, based on the self-declared origin of individuals as either Hispanic or not Hispanic. We subdivide Hispanics into two groups: those of Puerto Rican origin, and other Hispanics. Of course, non-Puerto Rican Hispanics are a heterogeneous group, consisting of Cubans, Mexican-Americans, Europeans, Central and South Americans and others. Thus, conclusions about the Hispanic group refer to the aggregate of such persons and do not necessarily apply equally to each group within this overall aggregate. "Black" refers to blacks who are not Hispanic. Persons who are neither black nor Hispanic are called "white non-Hispanics" or simply "whites." Note, however, that the group we call whites includes a relatively small number of Orientals, American Indians, and others who are not necessarily Caucasian.

By "labor force status" we mean the conventional trichotomy used in most government surveys modified so as to distinguish between employment in the federal sector and employment in the non-federal sector. Thus, in our analyses, any individual's labor force status is always one of the following mutually exclusive and exhaustive conditions: employed in the federal sector, employed in the non-federal sector, unemployed (that is, not employed but seeking employment), or not in the labor force.

Finally, by "ethnic wage discrimination" we mean any difference in total compensation—including both pecuniary and nonpecuniary compensation--that is associated with differences in ethnicity but is not associated with differences in productivity. This definition seems to be standard (for example, see Arrow, 1973, p. 4). Our definition emphasizes something that, while implicit in most definitions of wage discrimination, is worth noting explicitly: wage discrimination means differences in total compensation, rather than just in pecuniary compensation per se. For example, under our definition, pay differentials that are purely compensating or equalizing in nature are not discriminatory even if they are associated with ethnicity but not productivity. By the same token, the absence of a difference in pecuniary compensation may also entail wage discrimination. For example, an employer who offers Hispanic workers the same pecuniary pay but less desirable working conditions than equally productive non-Hispanic workers is behaving in a discriminatory manner, in our sense of that term.

This paper is organized as follows. We first present the economic theory underlying our statistical models, and then discuss the statistical models. We next present a summary of the data used, discuss our results regarding ethnic differences in labor force status, and describe the direct regression results from the Survey of Income and Education data. The reverse regression results from the SIE data follow; we then discuss the structural regression results from the same data. The next section discusses statistical results on federal compensation derived using an alternative data set, followed by comparison of all the statistical results. The final section presents our conclusions.

THE THEORETICAL MODEL

Like most branches of economics, labor economics is concerned with the analysis of supply and demand. As an actual or potential employee, the individual is chiefly concerned with the labor supply decision: he must decide how much to work and the sector in which to work subject to the constraints he faces. Thus, the individual is a constrained utility-maximizer, in the neoclassical sense: he selects the combination of work hours, leisure hours, and job characteristics (including both pecuniary and nonpecuniary compensation) that brings the highest possible level of happiness consistent with the constraints. Sometimes this maximum entails not working at all--for example, individuals who do not succeed in obtaining a job offer over a given period obviously will not be able to work, and other individuals may find that being in school or retirement is more desirable than employment--in which case the individual is either unemployed or not in the labor force. Since the individual maximizes subject to constraints, it makes sense to say that choices are voluntary only if one adds that they are made subject to whatever constraints exist.

While individuals, considered as agents in the labor market, are concerned with the labor supply decision, the major concern of the firm, as an actual or potential employer, is the labor demand decision. The firm must decide how high a wage it is willing to offer and what types of jobs it requires. Faced with a competitive market for hiring employees, firms do not offer more than is necessary to attract proper employees nor offer less than is necessary to fill all positions.

Firms may be viewed as continually making job offers, consisting of pecuniary compensation and a package of job characteristics which, in effect, constitute nonpecuniary compensation. Individuals may be viewed as continually seeking job offers and accepting or rejecting them. What is observed in a collection of data--for example, a sample survey--is the outcome of this job offer and job acceptance (or job rejection) process. The observed wage and employment outcome is the result of the process, not the process itself. For example, the fact that a given person selects a job in the federal sector over a job elsewhere is correctly called endogenous both to the individual's labor supply decision and to the labor demand decisions of employers.

An individual's sector of employment is at least partly a result of an economic decision by the individual about which job to accept (and about whether he will work at all). Each employer assesses the potential productivity of prospective emloyees by analyzing the skills they have to offer in light of the skills it needs. The employer offers prospective workers a package of pecuniary pay and other job characteristics intended to be attractive to them. At the same time, an individual who gets one or more offers decides whether to accept one (and, if so, which) or to reject all offers. After the decision, an outside analyst observes the resulting employment and unemployment. Observed differences in wages, job characteristics, or other outcomes (e.g., concentration of persons in a particular racial group in a particular sector) are all results of this process.

Since firms seek to maximize profits and understand that workers seek to maximize utility, firms will, on average, offer job packages consisting of pecuniary pay and working conditions that will fill the available positions at minimum cost. A firm whose offers are unnecessarily attractive will be flooded with applicants. It, and any competing enterprise, then knows that it can reduce the generosity of its offers, broadly defined, and still attract adequate numbers of applicants. Subject to some important qualifications to be noted below, the utility associated with a given job offer will then fall to the minimum level required to attract the number of workers the firm wants. In this way, then, firms rely on the nature of utility-maximizing behavior of individuals and on the nature of a competitive market to bring labor supply and labor demand into balance. In all cases, individuals decide which of the options available to them is best, subject to the constraints they face.

Of course, employers may sometimes decide, as a matter of conscious policy, to operate out of equilibrium, at least in the sense of an imbalance between the number of persons willing to work for the employer at the current level of generosity of the employer's job package (supply) and the number of positions the employer wants to fill (demand). For example, the federal sector may continually and deliberately make job offers with compensation in excess of the minimum necessary to fill the number of positions it wants to fill. This will result in a waiting list, or queue, for federal jobs. When such a queue exists, the various jobs available need to be allocated or rationed out among the applicants according to some method, formal or informal. For federal government employment, one such method of allocation is political--some of the available jobs may be allocated through a process of explicit or implicit payoffs. In this situation, different groups in the population have an incentive to compete for the political clout necessary for influence over the allocation process. The resources spent competing for such clout eventually bring the system back into equilibrium. If a federal job offers a premium over the minimum amount that the individual would require in order to be willing to accept it, then the individual will be willing to spend resources up to the amount of that premium to get enough clout to be offered that job.

Political allocation may help explain why the federal government can make better job offers and have higher minority employment relative to total employment than other employers. This higher relative minority employment may be in regions where minority political clout is higher. For example, minorities may have political clout in regions where minority population proportions are higher than they are in the country as a whole. This implies that measures of local population proportions for minorities may be relevant to analyses of federal employment.

Of course, non-federal employers, including employers in the private sector, may also--like the federal sector--make wage offers in excess of the minimum necessary to fill the number of positions they want to fill. Marginal private sector employers cannot do so because their profits would be driven below the minimum required for survival. Intramarginal private sector employers may do so if they choose. For example, a private sector employer with access to superior production technology will be more profitable than average; while this greater potential profitability may accrue to shareholders, it may instead take the form of wage offers to some groups that exceed the minimum required to fill the jobs the firm wants to fill. Similarly, a private sector employer may make unnecessarily high or excessive offers as a result of a collective bargaining agreement. In cases such as these, as in our previous discussion of job allocation through political clout, there will be a disequilibrium in the sense that, at the prevailing wage offer, defined broadly so as to include nonpecuniary as well as pecuniary rewards, supply will exceed demand. This will induce adjustments that will eventually bring the market back into equilibrium; as before, such adjustments involve expenditures of resources up to the amount of the premium implicit in the employer's offer. In some cases, such expenditures are implicit and occur through queueing. In other cases, such expenditures are explicit. In still other cases, supply and demand are equated through a rationing mechanism that has little to do with productivity considerations such as when the employer makes offers based on factors like race rather than on the basis of productivity.

The labor market, then, settles into an equilibrium in which the observed distribution of wages and the observed sectoral composition of employment are the result of demand and supply decisions. In what follows, we are concerned in general terms with intrasectoral differentials in employment and wage rates by ethnicity, with special reference to Puerto Ricans. To clarify the nature of some of the issues in which we are particularly interested, consider the following two questions:

Question 1: If one were to take a randomly selected group of individuals from the population of a given ethnic group and change their ethnicity to non-Hispanic (in the case of Hispanics) or to Hispanic (in the case of non-Hispanics), while keeping all of their measured and unmeasured productivity-related characteristics the same, then would the average of the wage offers made to such persons in a given sector differ from the offers that such employers would make if they knew the actual ethnicity of these individuals, and if so, by how much?

Question 2: If one were to take all the individuals in a given ethnic group who are employed in a given sector and change their ethnicity to non-Hispanic (in the case of Hispanics) or to Hispanic (in the case of non-Hispanics), while keeping all of their measured productivity-related characteristics the same, then would the average of their wages computed on the assumption that they were non-Hispanic (in the case of Hispanics) or Hispanic (in the case of non-Hispanics) differ from the actual average of their wages, and if so, by how much?

The answers to these two questions need not be identical. Both questions are of interest for most discussions of employer discrimination in the labor market. However, as we emphasize below, a particular statistical technique may provide a satisfactory answer to one of these questions without yielding any direct or useful evidence on the other.

STATISTICAL MODELS

Direct Wage Regression

The vast majority of studies of wage differentials by race, ethnicity, or sex rely on the methodology of direct wage regression. Under this procedure, one fits an earnings function--with a measure of pay such as earnings or wages as the dependent variable, and with measures of productivity-related characteristics and hypothetically irrelevant characteristics (sex, race) as independent variables--by applying least squares to data on individuals actually employed in some sector of interest. In some cases, as in Mincer's (1974) seminal work, sector means all employed persons. In other cases, sector refers to a single employer, as in the studies by Smith (1977), Malkiel and Malkiel (1973), Oaxaca (1976), Ehrenberg (1979), Osterman (1979), and many others. Regardless of how sector is defined, however, all such studies are investigating wages given that the individuals in the analysis are all in the sector being studied and have both received and accepted an offer from that sector.

It is important to understand what kind of evidence about the source and magnitude of wage and earnings differentials is contained in direct wage regression results. While direct wage regression may provide useful information on some questions, it may provide little or no direct evidence on others. Direct wage regressions analyze wage offers that have been received and accepted. Thus, while it appears that results derived from direct wage regressions may be quite useful for answering what we have called Question 2, they may be much less useful for answering what we have called Question 1.

At the statistical level, it is important to note that, considered only in terms of questions on which it can reasonably be expected to provide useful information, direct wage regression may provide evidence that is misleading--in particular, estimates that may be biased or inconsistent, in a statistical sense. Such bias or inconsistency can arise due either to exclusion of relevant variables or to inclusion of inappropriate variables. Inclusion of inappropriate variables--more generally, endogenous variables--such as occupation may bias direct wage regression results. Endogenous variables such as occupational status are dependent variables that, along with pay, are simply different aspects of the outcome of the interaction between supply and demand. Treating such variables as independent variables in a direct wage regression confuses cause and effect in a fundamental way.

Exclusion of relevant variables may also bias direct wage regression results. For example, prior occupational status may be regarded as a measure of the quality of one's work experience prior to becoming employed by one's present employer. It is therefore a productivity-related characteristic and, by definition, it is exogenous to the behavior of one's present employer. Omission of a potentially important productivity indicator of this kind may entail bias or inconsistency in the estimates of direct wage regression parameters.

The problem of omitted-variable bias has sometimes been misinterpreted or misunderstood, however. In particular, the fact that an omitted variable (e.g., prior work history or prior occupational status) is correlated both with the dependent variable and with an included independent variable does not mean that omission of the variable leads to bias in the coefficient of any particular independent variable included in the regression. Rather, a coefficient will be biased only if the omitted variable is correlated with the dependent variable and with the particularly independent variable at the margin, i.e., when all other independent variables are held constant. Thus, for example, in order to maintain that omission of prior occupational status will bias the coefficient on an ethnicity indicator variable, it is neither necessary nor sufficient to show that persons in different ethnic groups differ in terms of prior occupational status or that prior occupational status is associated with pay. Rather, one must show that persons in different ethnic groups with the same values for the included variables—age, educational attainment, and the like—nevertheless differ in terms of prior occupational status. Thus, the claim that the omission of variables that are plausibly associated with pay even at the margin inevitably biases the coefficient on an ethnicity variable in a direct earnings regression is not persuasive, even when there is reason to believe that persons in different ethnic groups differ in terms of such relevant omitted variables.

A different but related bias is induced by errors of measurement in the included variables. It would be surprising if such variables were always perfect surrogate or proxy measures of productivity, and it is possible that such variables measure actual or expected productivity with error. In this case the coefficients in a direct wage regression may be subject to what Roberts (1979, 1981) has called underadjustment bias. A statistical procedure used to address this problem is called reverse regression.

Reverse Wage Regression

The general phenomenon of measurement error bias in regression models has received attention for many years, and is a standard topic in many econometrics texts (e.g., Kmenta, 1971, pp. 307-322; Maddala, 1977, pp. 292–305). The problem of measurement error bias in direct wage regression, however, has received relatively little attention; most work on this subject is quite recent (e.g., Welch, 1973; Hashimoto and Kochin, 1979; Roberts, 1979, 1980, 1981; Kamalich and Polachek, 1982). Our discussion of measurement error bias in direct wage regression and the conditions under which reverse wage regression may avoid such bias will focus on the bivariate case: the relationship between pay and a single productivity-related characteristic. Either variable may be measured with error. (The analysis of the theory of reverse wage regression in the multivariate case involving the relationship between pay and a vector of productivity-related characteristics is much less tractable.)

Assume that the first two moments of the random variables y^*, p^*, e_1^*, and e_2^* are given by

$$(1)\quad E\begin{bmatrix} y^* \\ p^* \\ e_1^* \\ e_2^* \end{bmatrix} = \begin{bmatrix} \mu_y \\ \mu_p \\ 0 \\ 0 \end{bmatrix} = \underline{\mu},\ \mathrm{Var}[\cdot] = \begin{bmatrix} \omega_{11} & \omega_{12} & 0 & 0 \\ \omega_{12} & \omega_{22} & 0 & 0 \\ 0 & 0 & \omega_{33} & 0 \\ 0 & 0 & 0 & \omega_{44} \end{bmatrix} = \Omega,$$

where y^* is the appropriate pay variable, measured perfectly; p^* is the productivity index, measured perfectly; e_1^* is the measurement error in the pay variable; and e_2^* is the measurement error in the productivity variable. The observable pay, y, and observable productivity, p, are defined as:

$$(2a)\quad y = y^* + e_1^*$$

$$(2b)\quad p = p^* + e_2^*.$$

Accordingly, the first two moments of [y, p] are given by

$$(3)\quad E\begin{bmatrix} y \\ p \end{bmatrix} = \begin{bmatrix} \mu_y \\ \mu_p \end{bmatrix} \quad \mathrm{Var}[.] = \begin{bmatrix} \omega_{11} + \omega_{33} & \omega_{12} \\ \omega_{12} & \omega_{22} + \omega_{44} \end{bmatrix}.$$

The system described by equations (2)-(3) is a standard bivariate measurement error model. True pay, y^*, and true productivity, p^*, are subject to measurement errors e_1^* and e_2^*, respectively, which are assumed uncorrelated with each other and uncorrelated with the other variables in the true system. Since the measurement errors have zero expectation, the true variables, y^* and p^*, have the same expected values as the measured proxies, y and p, respectively. Since the measurement errors are uncorrelated with any other variables in the system, the measured proxies have the same covariance as the true variables. However, the variance of each measured variable exceeds the variance of its true counterpart by the variance of the measurement error.

We consider next the regression relationships connecting the true variables and the proxy variables. By definition, the regression of y^* on p^* can be decomposed into the conditional expectation of y^* given p^* and an expectation error which is uncorrelated with the conditional expectation. We will assume that the conditional expectations are linear in the conditioning variables. In addition, assume that the mean vector $\underline{\mu}$ and the system covariance matrix Ω are different for each race/ethnic group i, i = Hispanic, white non-Hispanic, and black non-Hispanic. For each race/ethnic group i, then, the regression relationships connecting the true variables are given by

(4a) $y^* = E[y^* \mid p^*]_i + \eta_1^*$

$= a_i^* + b_i^* p^* + \eta_1^*$

(4b) $p^* = E[p^* \mid y^*]_i + \eta_2^*$

$= \alpha_i^* + \beta_i^* y^* + \eta_2^*,$

where η_1^* and η_2^* are the errors of the conditional expectations and a_i, b_i, α_i, and β_i are the parameters of the linear functional form for the conditional expectations. When the true system is Multivariate Normal or the system is estimated by least squares using the true variables, the conditional expectation parameters are the following functions of the underlying system parameters:

(5a) $b_i^* = \frac{\omega_{12i}}{\omega_{22i}}$ $\qquad a_i^* = \mu_{yi} - b_i^* \mu_{pi}$

(5b) $\beta_i^* = \frac{\omega_{12i}}{\omega_{11i}}$ $\qquad \alpha_i^* = \mu_{pi} - \beta_i^* \mu_{yi}.$

When the true model is Multivariate Normal, these relationships hold exactly. When the true model is only specified up to its first two moments, as in equation (1), the relationships in (5) hold as the probability limits of the least squares estimators of the theoretical parameters when the true variables are used in the analysis.

Of course, only y and p are directly observable. Consequently, we must know the regression relationship connecting these variables in order to state the implications of the measurement error problem for the discrimination analysis of interest. The regression of y on p is defined as the conditional expectation of y given p. Once again, by the assumption of linear conditional expectations, the regression relationships connecting the observable variables for each race/ethnic group i are given by

$$\text{(6a)}\quad y = E[y^* \mid p^* + e_2^*]_i + \eta_1$$

$$= a_i + b_i\, p + \eta_1$$

$$\text{(6b)}\quad p = E[p^* \mid y^* + e_1^*]_i + \eta_2$$

$$= \alpha_i + \beta_i\, y + \eta_1.$$

When the true system (1) is Multivariate Normal or when the conditional expectations are estimated by least squares using the observed variables, the conditional expectation parameters in (6) have the following relationship to the theoretical parameters of the underlying system:

$$\text{(7a)}\quad b_i = \frac{\omega_{12i}}{\omega_{22i} + \omega_{44i}} \qquad a_i = \mu_{yi} - b_i\, \mu_{pi}$$

$$\text{(7b)}\quad \beta_i = \frac{\omega_{12i}}{\omega_{11i} + \omega_{33i}} \qquad \alpha_i = \mu_{pi} - \beta_i\, \mu_{yi}.$$

When the true model is Multivariate Normal, these relationships hold exactly. When the true model is only specified up to its first two moments, as in equation (1), the relationships in (7) hold as the probability limits of the least squares estimators of the theoretical parameters when the observed variables are used instead of the true variables. Notice that the presence of measurement errors e_1^* and e_2^* causes the theoretical regression parameters in equations (5)—the starred values—to deviate from the theoretical regression parameters in equation (7)—the unstarred values. Technically, the symmetric measurement error model has the property that the least squares estimators for the regression parameters a_i, b_i, α_i, and β_i are inconsistent estimators of the regression parameters a_i^*, b_i^*, α_i^*, and β_i^* connecting the true variables. However, it is straightforward to verify that the conditional expectation of the proxy pay variable given the true value of the productivity variable is identical to the conditional expectation in (4a). Similarly, the conditional expectation of the proxy productivity variable given the true pay variable is identical to the conditional expectation in (4b).

The inconsistency in the estimators based on the observed variables is at the heart of the criticisms leveled by Hashimoto and Kochin (1979) and Roberts (1979, 1981) against the direct regression methodology in statistical discrimination analyses. Direct regression is identical to least squares estimation of a_i and b_i. These estimators are inconsistent for the theoretical quantities a_i^* and b_i^* (or $\underline{\mu}_i$ and Ω_i). The effect of the inconsistency on the potential inference of statistical discrimination based on the direct regression estimates can be seen by considering the case in which each race/ethnic group has the same theoretical values of a_i^* and b_i^*. Then, the theoretical average difference in observed pay between a member of race/ethnic group i and a member of group j, conditional on the same true value of productivity, p*, is given by

$$(8) \quad E[y_i \mid p^*] - E[y_j \mid p^*] = a_i^* + b_i^* p^* - (a_j^* + b_j^* p^*) = 0,$$

since, by hypothesis, $a_i^* = a_j^*$ and $b_i^* = b_j^*$. However, if the least squares estimates of a_i and b_i are used, the estimated difference in pay between a member of race/ethnic group i and a member of group j, conditional on the same value of observed productivity, p, is given by

$$(9)\quad E[y_i \mid p] - E[y_j \mid p] = a_i + b_i\, p - (a_j + b_j\, p)$$

$$= a_i^* - a_j^* + (b_i^* - b_j^*)\, p + b_i^* \frac{\omega_{44i}}{\omega_{22i} + \omega_{44i}} \mu_{pi} - b_j^* \frac{\omega_{44j}}{\omega_{22j} + \omega_{44j}} \mu_{pj}$$

$$= b^* \frac{\omega_{44}}{\omega_{22} + \omega_{44}} (\mu_{pi} - \mu_{pj}),$$

since $a_i^* = a_j^*$ and $b_i^* = b_j^*$, by hypothesis. Notice that the expression in (9) is not necessarily zero unless $\mu_{pi} = \mu_{pj}$--that is, unless the average observed productivity index is the same for both groups. Normally, a test of the hypothesis of equal theoretical coefficients in the direct regression is considered a basis for an inference of statistical discrimination. Apparently, this test may support an inference of discrimination even though the theoretical coefficients of interest are equal when productivity is measured with error and the groups have different average values of the productivity proxy.

The analysis is symmetric in its implications for the reverse regression methodology. The least squares estimators of α_i and β_i are inconsistent for the theoretical parameters α_i^* and β_i^*. Reverse regression is identical to least squares estimation of α_i and β_i. The effect of the inconsistency on the potential inference of discrimination based on the reverse regression estimates can be seen by considering the case in which each race/ethnic group has the same theoretical values of α_i^* and β_i^*. Then, the theoretical average difference in observed productivity between a member of race/ethnic group i and a member of group j, conditional on the same true value of pay, y*, is given by

$$\text{(10)}\quad E[p_i \mid y^*] - E[p_j \mid y^*] = \alpha_i^* + \beta_i^* y^* - (\alpha_j^* + \beta_j^* y^*) = 0,$$

since, by hypothesis, $\alpha_i^* = \alpha_j^*$ and $\beta_i^* = \beta_j^*$. However, if the least squares estimates of α_i and β_i are used, the estimated difference in pay between a member of the race/ethnic group i and a member of group j, conditional on the same value of observed productivity, p, is given by

$$\text{(11)}\quad E[p_i \mid y] - E[p_j \mid y] = \alpha_i + \beta_i y^* - (\alpha_j + \beta_j y^*)$$

$$= \alpha_i^* - \alpha_j^* + (\beta_i^* - \beta_j^*) y + \beta_i^* \frac{\omega_{33i}}{\omega_{11i} + \omega_{33i}} \mu_{yi} - \beta_j^* \frac{\omega_{33j}}{\omega_{11j} + \omega_{33j}} \mu_{yj}$$

$$= \beta^* \frac{\omega_{33}}{\omega_{11} + \omega_{33}} (\mu_{yi} - \mu_{yj}),$$

since $\alpha_i^* = \alpha_j^*$ and $\beta_i^* = \beta_j^*$, by hypothesis. As we noted for expression (9), the mean difference in equation (11) is not necessarily zero unless $\mu_{yi} = \mu_{yj}$—that is, unless the average observed pay is the same for both groups. Apparently, the reverse regression also may support an inference of statistical discrimination even though the theoretical coefficients of interest are equal.

Although equations (9) and (11) are symmetric in their implications for the type of inconsistency induced by least squares analysis of the system (1) when only the system (2) is observed, the two inconsistencies lead to quite different errors in a statistical discrimination analysis. In general, the covariance between pay and productivity is positive

($\omega_{12} > 0$). Therefore, the estimated regression slope parameter is expected to be positive whether one estimates b^*, β^*, b, or β. Consequently, the sign of the inconsistency depends on the sign of the difference in the mean values of productivity or pay for each race/ethnic group. If ethnic group i has a higher value of the observed productivity index than ethnic group j, then equation (9) implies that direct regression analysis of the observable variables y and p will be biased in the direction of finding discrimination favoring group i even when all coefficients of interest are equal. However, if race/ethnic group i has a higher mean value of observed pay than race/ethnic group j, then equation (11) implies that reverse regression analysis of the observable variables will be biased in the direction of finding discrimination favoring group j even when all coefficients of interest are equal.

Roberts (1981) has called this phenomenon the conflict between two potential definitions of statistical discrimination. Under his first definition, differences in true pay, y^*, given the same values of true productivity, p^*, are evidence of statistical discrimination: that is, a racial/ethnic group is discriminated against if it has lower expected true pay for a given level of true productivity. As Roberts notes, and equation (9) shows, direct regression estimation of the conditional expectation of observed y given observed p may give spurious evidence of statistical discrimination in the case where one group simply has a higher average value of the productivity proxy p than the other. Under Roberts' second definition of statistical discrimination, differences in true productivity, p^*, given the same values of true pay, y^* are evidence of discrimination: that is, a racial/ethnic group is discriminated against if it has a higher expected true productivity for a given level of true pay. As Roberts notes, and equation (11) shows, reverse regression estimation of the conditional expectation of observed p given observed y may also give spurious evidence of statistical discrimination in the case where one group simply has higher average measured pay y than the other group. In principle, however, the errors involved in using direct or reverse regression are in the opposite direction. That is, if the observed average pay of group i is greater than the observed average pay of group j, then the observed average productivity of group i is very likely to be higher than the observed average productivity of group j. Under these conditions, direct regression analysis of the proxy variables y and p may lead to an inference of discrimination against group j while reverse regression analysis of the same proxy data may lead to an inference of discrimination against group i.

The direct and reverse conditional expectation definitions of statistical discrimination are not actually different. When applied to the true variables y^* and p^*, either definition of discrimination leads to the same implications for the structural parameters $\underline{\mu}$ and Ω, as equations (8) and (10) show. In general, true pay cannot be measured exactly since the appropriate measure would include current compensation, fringe benefits, the monetary value of future promotion possibilities, future benefits, and on-the-job amenities. Similarly, true productivity cannot be measured exactly since the true index depends on schooling, types and quantities of previous experience, and various other factors that may be difficult to quantify. The importance of the analysis of direct and reverse regression methods for estimating the parameters underlying either definition of statistical discrimination is that, under typical conditions, the two statistical methods will result in estimates that bound the actual magnitude of discrimination. (However, as noted below, a potential problem with either direct or reverse methodology is the implicit assumption that, if the structure in equation (1) differs across race/ethnic groups in such a way that either equation (8) or (10) is not zero, then such structural differences can erroneously be interpreted as differences in the behavioral equations governing the employment practices of the employer or sector being analyzed.)

We have derived a version of the reverse wage regression method for use in analyses comparable to the direct regression models. The procedure involves two steps. In the first or "direct" stage, we compute an underlying direct regression using a randomly selected half of the white non-Hispanic observations available to us. We use only half of the available observations to fit the direct regression because these estimated coefficients will be used to form a productivity index for the remaining half of the white non-Hispanics and all the black and Hispanic observations. (Splitting the sample avoids inducing spurious correlation between the computed productivity index and the wage rates in the reverse regressions.) The direct regressions used in the first stage involve all the productivity indicators used in the direct regression except, of course, the ethnicity indicators and interactions involving these indicators.

In the second or "reverse" stage, we use the conventional wage or earnings function coefficient estimates from the direct stage to compute predicted wages or earnings y for the remaining observations. We treat this constructed variable $\tilde{y}$ as a proxy measure of productivity. Accordingly, $\tilde{y}$ becomes the dependent variable in our second-stage reverse wage regression. We compute

$$(12) \quad \tilde{y} = a + \underline{b_0}'\underline{d} + b_1 y + \eta,$$

where $\underline{d}$ is a vector of race and ethnicity indicators, y is a measure of pay (e.g., the logarithm of the hourly wage), and η is the regression error term. Thus $\tilde{y}$ is a linear function of y (and $\underline{d}$).

Structural Wage Regression

Both direct and reverse wage regression are concerned with conditional wage relationships. Such techniques are therefore directly concerned with what we have called Question 2—identifying the within-sector differences in wages and earnings for different race/ethnic groups. However, they do not, in general, estimate the parameters governing the structure of the underlying process of supply and demand that generates wage offers; rather, they constitute analyses of the outcome of that process. Neither direct nor reverse wage regression addresses what we have called Question 1—identifying the across-sector differences in wages and earnings opportunities for different race/ethnic groups.

In order to obtain answers to Question 1, it is necessary to address directly the question of the determinants of wage offers. Unfortunately, most data sets, particularly survey data sets, contain information on only a subset of all wage offers—namely, the ones that have been both received and accepted. In particular, in terms of our federal/non-federal sector dichotomy, most cross-sectional survey data on any given individual contain information on only one offer (from either the federal or the non-federal sector) for employed persons, and do not contain information on any offer, from either sector, for persons who are unemployed or not in the labor force.

Such data are said to be censored, in the sense that the investigator does not know the values of certain variables of interest: in the present case, he does not know the values of the federal sector offers available to persons working in the non-federal sector or the values of non-federal sector offers available to persons working in the federal sector; moreover, he does not know the values of the offers from either sector that are available to persons who are unemployed or not in the labor force. Restricting one's analysis to a given sector aggravates the problem: intrasectoral data are truncated, in the sense that a sample consisting exclusively of intrasectoral data is one from which data on persons outside the sector being analyzed have been discarded.

To ignore this truncation completely, as in an intrasectoral direct or reverse wage regression analysis, may subject a study to sample selection bias, at least insofar as answers to Question 1 are concerned (see Heckman, 1979; Heckman, Killingsworth, and MaCurdy, 1981). Sample selection bias may arise in such a study because the data to be used contain only observations on persons who have received and accepted an offer from the sector in question. For example, the observations contained in data for a given sector are in part self-selected, in the sense that, having received an offer from employers in that sector, the persons observed in the data for that sector have all selected themselves into the sample to be analyzed. Application of direct or reverse wage regression to a self-selected sample of this kind may not yield consistent estimates of the parameters of the employer's wage offer function. More generally, a sample of this kind has a sampling distribution determined by both the survey design and the respondent in the sense that it consists of persons who have accepted offers. This makes it not only a self-selected sample, in the sense used above, but also a "selected sample" in the sense that such persons must first have received offers from, and thus must have been selected by, employers.

This suggests that one way to avoid the self-selection biases that may arise in the context of direct or reverse regression analysis of an intrasectoral sample is to derive a model that not only (i) specifies the determinants of wage offers--the relation of primary interest--but also (ii) describes the process of selection by which the individuals in such a sample got into the sample. We start by deriving a model of the selection process, and then show how this model may be used in conjunction with a model of the determinants of wage offers to obtain consistent estimates of the structural wage offer function.

Since the data in the 1976 Survey of Income and Education (SIE), which are used in most of the studies discussed here, refer to a period of unusually severe recession, it is worth noting that problems associated with selection bias may be more important in these data than they would be in data that referred to a period when business-cycle conditions were more normal. For example, results based on direct (or reverse) wage regression analyses of these data might lead to misleading inferences about employer offers by virtue of the fact that nonemployment--either unemployment or absence from the labor force induced by the 1975-76 downturn--during 1975-76 was well above the level observed in more normal periods. In constrast, structural regression in effect makes a statistical correction for possible biases that might be introduced by such pheonomena. Employer wage offers may themselves be affected by cyclical downturns such as the one observed during 1975-76, and structural regression techniques cannot be used to correct for the impact of a slump on wage offers as such. However, structural regression techniques do at least permit a correction for the way in which a cyclical downturn—and the rise in nonemployment during a downturn—might otherwise confound attempts to obtain unbiased measures of the determinants of employer wage offers.

We first derive a model of the way in which individuals are selected into different sectors—i.e., of the determinants of the labor force status of individuals, categorized, as before, as being (i) employed in the federal sector, (ii) employed in the non-federal sector, (iii) unemployed, or (iv) not in the labor force. This model may be used to compute labor force status probabilities (i.e., the probability that labor force status will be any one of these four distinct categories) for every individual. These probabilities may then be used to form instrumental variables for structural wage regression.

The basic notion underlying our model of labor force status determination is the idea of an index function model (see Heckman, Killingsworth, and MaCurdy, 1981) or, more or less equivalently, a discrete choice model (see McFadden, 1973, 1975). An index function model represents the decision-making process of an agent who is faced with the problem of having to choose the best of several alternatives. Associated with each alternative is a particular payoff or reward that is represented by the value of an index. The alternative actually chosen is the one with the highest index--that is, the one with the biggest payoff.

Specifically, recall that we have established four alternative possibilities for labor force status, and let the utility or payoff U associated with each possibility, or sector, s, be given by

$$(13)\quad U_s = V(w_s, q_s, \underline{x}) + v(w_s, q_s, \underline{x}),$$

where V, the systematic component of U, is a function of the wage offered to the individual by employers in that sector; q_s is an index of the characteristics associated with that sector (e.g., one's home or school environment, for the "not in the labor force" sector; the work environment, for the federal employment sector); $\underline{x}$ is a vector of observed characteristics of the individual; and v is an error term (the stochastic component of U). Note that no wage is relevant to being in the unemployed sector or the "not in the labor force" sector. The individual will choose to be in a particular sector s if the utility associated with that choice exceeds the utility associated with any other choice. For example, the individual will choose the federal sector if and only if

$$(14)\quad U_f > \mathrm{Max}(U_n, U_u, U_o),$$

where the f subscript refers to the federal sector, n refers to the non-federal sector, u refers to the unemployment sector, and o refers to the "not in the labor force" sector. Expressions similar to (14) define the circumstances under which the individual willl choose non-federal employment, unemployment, or absence from the labor force. Note that all such choices are subject to the values of the wage offers received from the federal and non-federal sectors, w_f and w_n. Thus, as before, choice is subject to constraints, and statements that choice is voluntary make sense only if one understands both that such choices are constrained and, thus, that the fact that such choices are voluntary has no particular normative implications. Note also that non-receipt of an offer from the federal or non-federal sector may be treated as, and is treated in this analysis as, the equivalent of receipt of a very low offer from that sector.

To specify the decisions process (13)-(14) in a manner suitable for empirical estimation, let the systematic component V of the utility function for sector s (s = f, n, u, or o) be given by

$$(15) \quad V(w_s, q_s, \underline{x}) = a_1(q_s)\, w_s + \underline{x}'\underline{a_2}(q_s),$$

where $a_1(.)$ and $\underline{a_2}(.)$ are, respectively, a scalar and a vector function of q_s, which vary across sectors because of their dependence on the characteristics q_s of that sector. Next, assume that the logarithm of the (best) wage offer available to the individual from employers in sector s (s = f or n) is given by

$$(16) \quad w_s = \underline{z}'\underline{b_s} + e_s,$$

where $\underline{z}$ is a vector of observed variables that affect the wage offer w_s and e_s is an error term whose population mean is zero. Substitute (16) into (15) and rearrange terms, to obtain

$$(17) \quad U_s = \underline{z}'\underline{\gamma}_{1s} + \underline{x}'\underline{\gamma}_{2s} + v_s^* \underline{=} V_s^* + v_s^*,$$

where $\underline{\gamma}_{1s} = \underline{b}_s\, a_1(q_s)$

$\underline{\gamma}_{2s} = \underline{a_2}(q_s)$

$v_s^* = e_s\, a_1(q_s) + v(w_s, q_s, x)$,

which is linear in all observed variables $\underline{z}$ and $\underline{x}$. (Note that some elements in $\underline{z}$ may also appear in $\underline{x}$, and vice versa).

Finally, let the distribution of the random term v_s^* in (17) be approximately independent Weibull. This means that intersectoral differences between these errors, $v_f^* - v_n^*$, $v_f^* - v_u^*$, $v_f^* - v_o^*$, etc., are all approximately independent logistic.

Together with (14), the independent logistic assumption implies that

$$(18)\quad \Pr\{\text{in sector } s\} = \frac{\exp(V_s^*)}{\exp(V_f^*) + \exp(V_n^*) + \exp(V_u^*) + \exp(V_o^*)}$$

for s = f, n, u, or o. Thus, (18) gives the probability that an individual will be in any given sector s as a logistic function of $\underline{x}$ and $\underline{z}$. Note that (18) is therefore a reduced form expression, since it contains both supply and demand variables.

We now consider how to use estimates of parameters governing labor force status, i.e., estimates of (18), to obtain estimates of the parameters of the wage offer equation. We refer to this as structural wage regressions.

As noted earlier, we consider two kinds of employment in our analyses: federal and non-federal employment. Let N_s be the number of persons in sector s; s = f or n. Let w_s be the logarithm of the (best) wage offer for work in sector s available to an individual with characteristics $\underline{x}$, $\underline{z}$, and assume that w_s is given by (16) above.

Now, (16) is an expression for the wage w_s that the individual will receive if he works in sector s and, by assumption, the mean value of w_s in the population as a whole, given $\underline{z}$, is

$$(19) \quad E[w_s \mid \underline{z}] = \underline{z}'\underline{b_s}.$$

On the other hand, the mean value of w_s, given $\underline{z}$, among persons actually working in sector t is

$$(20) \quad E[w_s \mid \underline{z},\ s = t] = \underline{z}'\underline{b_t} + E[e_s \mid \underline{z},\ s = t].$$

Note that (19) and (20) are equivalent only if the conditional mean of e_s is independent of the condition s = t, i.e., only if the population mean of the error term e_s and the mean of e_s among persons actually employed in sector t are the same. If not, then, in terms of the discussion in the previous section, persons in sector s are a selected sample. The sampling distribution of the e_s in the data is not the same as the distribution of the e_s in nature. This is the case in which conventional least squares analysis of the regression based on a sample restricted to persons actually in sector s will yield biased estimates of the parameters of the wage offer function b_s. Such a regression in effect ignores the second term on the right-hand side of (20), and so will suffer from omitted variable bias, where the omitted variable in question is the conditional mean of e_s. (For further discussion of this point, see Heckman, 1979.)

To derive an alternative to conventional regression that may be used to obtain consistent estimates of the parameters of the wage offer function, note that

$$(21)\quad E[w_s \mid \underline{z},\ s = f] =$$

$$\frac{\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} w_f\ \pi_f(w_f,\ w_n,\ \underline{x})\ p(w_f,\ w_n \mid \underline{z})\ dw_f\ dw_n}{\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \pi_f(w_f,\ w_n,\ \underline{x})\ p(w_f,\ w_n \mid \underline{z})\ dw_f\ dw_n},$$

where $\pi_s(w_f,\ w_n,\ \underline{x}) = \Pr\{\text{in sector } s \mid w_f,\ w_n,\ \underline{x}\}$ and $p(w_f,\ w_n \mid \underline{z})$ = the joint density function of w_f, w_n conditional on $\underline{z}$. Approximate the numerator of (21) with a first order Taylor series around the means of w_f and w_n. Approximate the denominator of (21) with the unconditional probability of choosing sector s to obtain an overall approximation:

$$(22)\quad E[w_s \mid \underline{z},\ s = t] = \underline{z}'\underline{b_t}\ \frac{\pi_t(\underline{z}'\underline{b_f},\ \underline{z}'\underline{b_n},\ \underline{x})}{\bar{\pi}_t},$$

where $\pi_s(w_f,\ w_n,\ x)$ has been evaluated at mean values of w_f and w_n, and $\bar{\pi}_s$ is the average value of π_s in the population. Note that π_s is the probability that an individual will be in sector s and may be computed using estimates of the parameters of (18), while $\bar{\pi}_s$ is the proportion of all persons in sector s.

Equation (22) suggests an instrumental variable estimator of the coefficients $\underline{b_s}$ in the structural wage equation (16). The basis for this claim is the form of the approximation to the conditional expectation of the wage given the sector of employment in equation (22). This is the approximate regression function for w_s given employment in sector s and the exogenous variables $\underline{z}$. Therefore, by construction, the variables on the right-hand side of (22) are orthogonal to the error term in the sector-specific wage regression. These right-hand side variables depend on an unknown ratio $\lambda = \pi(\underline{z}'\underline{b_f}, \underline{z}'\underline{b_n}, x)/\overline{\pi}_s$, which is the ratio of the probability of being employed in sector s evaluated at the mean value of the wage in each sector, given $\underline{z}$, to the average probability of being employed in sector s. This ratio fluctuates around unity. It is higher for individuals with higher than average probabilities of being in sector s and lower for individuals with lower than average probabilities of being in sector s. This ratio may be estimated by using as the numerator probability the fitted value of the estimated logit probability developed above and using as the denominator probability the sample proportion in sector s.

Having developed an estimator for this ratio, we are faced with a choice of strategies for estimating $\underline{b_s}$. First, we could regress the sector-specific wages on the product of $\underline{z}$ and the ratio λ. Since the ratio λ is estimated, this strategy will lead to problems in determining the appropriate measure of precision for this estimator. Alternatively, one may use λ to develop a set of instruments that are correlated with $\underline{z}$ but uncorrelated with the error in the conditional wage expectation given $\underline{z}$ and the sector of employment. These instruments are exactly the right-hand side of equation (22). The λ must still be estimated; however, this approach does not lead to problems in estimating standard errors because the convergence of the moment matrix of the instruments is guaranteed by the consistency of the logit parameter estimates and by the fact that no nonlinear instruments are used as right-hand side variables in the equation being estimated. The estimated residuals may be heteroskedastic; however, in estimation we allow for this possibility.

Each row of the instrument matrix Q is defined as

$$(23)\quad \underline{q}_{1i} = \underline{z}_i \lambda_{si},$$

where $\lambda_{si} = \pi(\underline{z}_i{}'\underline{b}_f, \underline{z}_i{}'\underline{b}_n, x_i)/\overline{\pi}_s$, $i = 1, \ldots, N_s$, and N_s = the total sample in sector s. To allow for potential misspecification of the probability-generating process we add a set of instruments, q_{2i}, defined as

$$(24)\quad \underline{q}_{2i} = \underline{z}_i \lambda_{si}^2.$$

The complete instrument matrix Q, then, consists of N_s rows of $[\underline{q}_{1i}{}', \underline{q}_{2i}{}']$. The b_s are estimated using instrumental variables:

$$(25)\quad \underline{\hat{b}_s} = [Z_s{}'Q_s(Q_s{}'Q_s)^{-1}Q_s{}'Z_s]^{-1}Z_s{}'Q_s(Q_s{}'Q_s)^{-1}Q_s{}'\underline{w}_s,$$

where Z is the N_s by k matrix of wage equation variables, Q is the N_s by 2k matrix of instruments, and $\underline{w}_s$ is the N_s by 1 vector of wages observed in sector s. The estimator of the asymptotic variance-covariance matrix is

$$(26)\quad \mathrm{Var}[\underline{\hat{b}_s}] = \hat{\sigma}_s^2\,[Z_s{}'Q_s(Q_s{}'Q_s)^{-1}Q_s{}'Z_s]^{-1},$$

where $\hat{\sigma}_s^2$ is the sum of squared structural residuals divided by the sample size N_s

$$(27)\quad \hat{\sigma}_s^2 = (\underline{w} - Z'\underline{\hat{b}_s})'(\underline{w} - Z'\underline{\hat{b}_s})/N_s.$$

Conceptually, the structural estimator of the parameters relating w_s to $\underline{z}$ is quite different from both the direct and reverse regression estimators of those parameters. If $\underline{z}$ includes a vector of race/ethnic indicators, say d, then the structural model developed in this paper estimates the coefficients on $\underline{d}$ for the population conditional expectation of w_s given $\underline{z}$ and not for the subpopulation conditional expectation of w_s given $\underline{z}$ and $s = t$ for some sector t. This difference is important, since the structural model attributes behavioral significance to the population conditional expectation and not to the self-selected subpopulation conditional expectation. In a direct or reverse regression analysis of race/ethnic pay differences, the conditional expectation of pay, given the productivity index $\underline{z'b_s}$ and given the sector of employment s, may differ across groups because of systematic differences in the employers' pay practices (the usual assumption in statistical discrimination analyses) or because of systematic differences in the workers' preferences, as modeled by the sectoral choice model above. In general the conditional expectation of pay, given $\underline{z'b_s}$ and sector s, may differ across race/ethnic groups because of variation in labor demand (employer policies) or labor supply (employee policies). The structural model developed in this chapter makes assumptions sufficient to identify the parameters underlying labor demand (but not labor supply), permitting estimation of the conditional expectation of pay offers given only $\underline{z}$.

DATA USED IN EMPIRICAL STUDIES

Most of the data used in the empirical studies described in this report are derived from the 1976 Survey of Income and Education (SIE).

The SIE was conducted during April-July 1976 by the U.S. Bureau of the Census for the U.S. Department of Health, Education and Welfare (now the Department of Health and Human Services), and was the largest national survey since the 1970 Census of Population. Most of the procedures and definitions used in the SIE are identical to those used in the annual March Current Population Survey (CPS), but the SIE also contains questions pertaining to income, education, and language skills that are not contained in the CPS. For further description of the SIE, see U.S. Bureau of the Census (1978).

We have excluded persons under 21 years old. Among persons 21 years old or older in the SIE data base, 8,168 are Hispanics, 19,501 are black non-Hispanics, and the remainder (246,837) are whites (that is, persons neither Hispanic nor black). Ethnicity is self-reported. Race, however, is determined by interviewer observation.

Second, we have excluded persons not residing in the continental United States; our data therefore exclude persons residing in Hawaii and Alaska, and also, of course, persons living in Puerto Rico.

The SIE therefore refers to a sample of persons in the country as a whole, and geography undoubtedly has major effects on pay through its association with such factors as (i) regional cost-of-living differentials, (ii) regional differences in amenities and also, to the extent that labor is immobile, (iii) regional differences in factor proportions (for example, see Kiefer and Smith, 1977). Moreover, there are important regional differences in the location of minority populations and the location of various industries, including the federal government. In all of our analyses, geography, specifically locational choice, is taken as exogenous. Nevertheless, we have taken several measures to ensure that minority groups are compared with nonminority groups from the same geographic region. The sampling design of the SIE oversampled less populated states, meaning that the geographic distribution of employment opportunities is not sampled randomly.

In order to control for the differences in labor demand across geographic regions, we have used two sets of geographically matched samples in our analyses. The logit models of the labor force status were estimated using samples of blacks and of white non-Hispanics that were geographically matched to our sample of Hispanics. Regression analyses were performed on federal and non-federal samples that were geographically matched to the federal sample.

We did this geographic matching by state and by what the SIE calls central city code, which categorizes persons according to residence in the following way: (1) located in the central city of a Standard Metropolitan Statistical Area (SMSA), (2) located in an SMSA but not in a central city, (3) located outside an SMSA, and (4) location not disclosed (in order to avoid breaching Census regulations governing confidentiality). Relatively small numbers of persons, mainly persons residing in outlying areas, fall into the last of these four categories. Thus, for example, after determining the total number of Hispanics living in the central city area of the Los Angeles-Long Beach SMSA in California, we randomly selected equal numbers of black non-Hispanics and of whites from the total populations of such persons in the same area; and similarly for all other areas. The result of this process of matching was three samples (of Hispanics, black non-Hispanics, and whites, respectively) with the same sampling probabilities for each state and central city code. In addition, we produced two samples (of federal and non-federal employees, respectively) with the same sampling probabilities for each state and central city code. Therefore, five analysis samples were produced: three which were geographicaly matched to the Hispanic data and two which were geographically matched to the federal data.

For the samples geographically matched to the Hispanic sample from the SIE, the sampling probabilities for Hispanics and whites are identical for each state and central city code. However, because there were not enough black non-Hispanics in the original SIE sample for the West and Southwest regions, this group is undersampled for these regions in our sample. All federal employees in the SIE are included in the federal sample. In the non-federal sample, whites are exactly matched geographically but Hispanics and black non-Hispanics are oversampled. Since ethnicity and location are always conditioning variables in the analyses using the federal and non-federal samples, the oversampling of blacks and Hispanics can be expected to reduce sampling error on ethnicity effects without inducing a location bias.

Since we are not able to observe the actual work experience of the individuals in our data, we must use a measure of potential work experience (Mincer, 1974) defined as current age less years of schooling less 5. The problems associated with this proxy are well known, particularly as regards male-female differences in potential vs. actual work experience. Accordingly, we think it appropriate in analyzing differentials in employment status, wages, and earnings to consider men and women separately.

Annual earnings, as defined in our studies, is the total amount of income from work received during the year 1975. The hourly wage, as used in our studies, is computed as the ratio of annual earnings to annual hours of work, where the latter is computed as the product of weeks worked during the year 1975 and usual hours worked per week during the year 1975. Labor force status is defined according to standard Current Population Survey concepts (U.S. Bureau of the Census, 1978) as of the week preceding the actual survey date.

The period 1975-76 was part of an unusually severe recession. This may have implications for the interpretation of our results. In particular, differentials of any kind (skill, racial, etc.) may tend to widen during business-cycle slumps and narrow during booms. To the extent that this is true, the various effects we discuss in this report may overstate somewhat the effects that would be observed during more normal (less recessionary) times.

In addition to the SIE we also used the federal government's Central Personnel Data File (CPDF). The CPDF is a payroll data set based on federal personnel files. CPDF data are derived from various federal payroll documents and are used by the Federal Office of Personnel Management and other federal agencies in studying characteristics of the federal civilian work force, in personnel planning, and in other related activities. The CPDF is longitudinal in nature, having begun in 1972 and having been updated on an annual basis since that time; thus, it permits analyses of several different years. Finally, since the CPDF covers essentially all federal employees, it contains large numbers of Hispanics as well as large numbers of persons in other racial and ethnic groups. (For further details on the CPDF, see Schneider, 1974.)

In computing results using the CPDF, we started with samples of 5,000 Hispanics and 5,000 non-Hispanics, selected randomly from the total CPDF populations present in each of the years 1975, 1976, and 1977. As in our work on the SIE data, we then excluded persons who either (i) were not living in the continental United States or (ii) were under 21 years old. This reduced a given year's sample by about 12% to about 8,800 people. About 15% of the persons remaining in any given year's sample after application of this exclusion could not be included in the regression for that year due to missing data (mainly for educational attainment or, to a lesser extent, race or sex). Also, we computed regressions for each year separately for each sex. Thus, the total size of the sample used for regressions for a given sex for a particular year is between about 2,000 (in the regressions for women) and about 5,600 (in the regressions for men).

In order to provide a basis for comparisons between the various statistical procedures described earlier, we estimated a set of different wage and earnings models using the same data and definitions. We briefly discuss the design of these models. All regression models for wages and earnings based on the SIE use the same sets of explanatory variables. The regression models for wages and earnings based on the CPDF use different but similar explanatory variables. The logit models for employment sector based on the SIE use an abbreviated set of explanatory variables. We describe each explanatory variable list in turn.

The dependent variable for the wage and earnings analyses based on the SIE is either the log of the hourly wage rate or the log of annual earnings. Independent variables capture effects on wages associated with human capital, ethnicity, race, age, geography, and other factors. A list of all variables used in the wage and earnings regressions based on the SIE data is as follows:

Dependent Variables

either the natural logarithm of the hourly wage rate
or the natural logarithm of annual earnings

Independent Variables

Group A variables (ethnicity and race indictors—variant 1):

1 if Hispanic, 0 otherwise

1 if black and not Hispanic, 0 otherwise

Group B variables (ethnicity and race indicators—variant 2):

1 if Puerto Rican, 0 otherwise

1 if Hispanic but not Puerto Rican, 0 otherwise

1 if black and not Hispanic, 0 otherwise

Group C variables (human capital, geography, and other factors):

number of years of formal education

1 if graduated from high school, 0 otherwise

1 if graduated from college, 0 otherwise

1 if any postgraduate education, 0 otherwise

1 if currently a full-time student, 0 otherwise

1 if currently a full-time public school student, 0 otherwise

number of years of education received outside the U.S.

1 if had any education outside the U.S., 0 otherwise

1 if taught in English, 0 if taught in any other language

1 if U.S.-born, spoke English as a child, and speaks English now; 0 otherwise

1 if not U.S.-born, 0 otherwise

number of years lived in U.S. (equal to zero, for persons born in U.S.)

1 if English not the primary language spoken as a child, 0 otherwise

1 if English not the primary language spoken now, 0 otherwise

1 if English not spoken or understood very well, 0 otherwise

1 if has any physical condition limiting ability to work, 0 otherwise

1 if age is over 30 and under 41, 0 otherwise

1 if age is over 40 and under 51, 0 otherwise

1 if age is over 50 and under 65, 0 otherwise

1 if age is over 64, 0 otherwise

potential experience (age minus years of schooling minus 5)

square of potential experience

1 if employed part-time, 0 otherwise

1 if a veteran, 0 otherwise

1 if lives in New England area (Maine, New Hampshire, Vermont, Massachusetts, Rhode Island, Connecticut), 0 otherwise

1 if lives in Middle Atlantic area (New York, New Jersey, Pennsylvania), 0 otherwise

1 if lives in East North Central area (Ohio, Indiana, Illinois, Michigan, Wisconsin), 0 otherwise

1 if lives in West North Central area (Minnesota, Iowa, Missouri, North Dakota, South Dakota, Nebraska, Kansas), 0 otherwise

1 if lives in South Atlantic area (Delaware, Maryland, District of Columbia, Virginia, West Virginia, North Carolina, South Carolina, Georgia, Florida), 0 otherwise

1 if lives in East South Central area (Kentucky, Tennessee, Alabama, Mississippi, Arkansas, Louisiana, Oklahoma, Texas), 0 otherwise

1 if lives in Pacific area (Washington, Oregon, or California), 0 otherwise

Group D variables (population proportions and interactions):

proportion of population in area (classified by state, SMSA, and central city) that is black non-Hispanic

proportion of population in area that is Hispanic

proportion black non-Hispanic in area times years of school

proportion Hispanic in area times years of school

proportion black non-Hispanic in area times potential experience

proportion Hispanic in area times potential experience

Group E variables (interactions with race, ethnicity indicators):

Hispanic indicator times years of school

black non-Hispanic indicator times years of school

Hispanic indicator times high school graduation indicator

black non-Hispanic indicator times high school graduation indicator

Hispanic indicator times college graduation indicator

black non-Hispanic indicator times college graduation indicator

Hispanic indicator times postgraduate education indicator

black non-Hispanic indicator times postgraduate education indicator

Hispanic indicator times potential experience

black non-Hispanic indicator times potential experience

Hispanic indicator times square of potential experience

black non-Hispanic indicator times square of potential experience

Group F variables (interactions between race, ethnicity indicators, and population proportions):

black non-Hispanic indicator times percent black non-Hispanic in area

black non-Hispanic indicator times percent black non-Hispanic in area times years in school black non-Hispanic indicator times percent black non-Hispanic in area times potential experience

Hispanic indicator times percent Hispanic in area

Hispanic indicator times percent Hispanic in area times years in school

Hispanic indicator times percent Hispanic in area times potential experience

Group A and Group B variables are indicators for minority status. Group A identifies Hispanics and blacks who are not Hispanics. Group B uses the same black non-Hispanic indicator but distinguishes between Hispanic subgroups, i.e., those of Puerto Rican origin and other Hispanics.

Group C variables are forms of the basic human capital variables normally found in direct wage regressions. The exact form of these variables is, of course, limited by the nature of the data available in the SIE. These variables—for education, age, potential work experience, and the like—are proxies intended to capture the employer's attempt to estimate the productivity of potential employees.

Some variables in Group C go beyond the basic proxies used in most previous research. Variables for years of education outside the United States and for not speaking English as one's primary language are intended to capture effects of immigration and language skills that may affect earnings (see Chiswick, 1978, 1980). Indicators of geographic location reflect the possible impact of region (that is, regional price differentials, capital-labor ratios, etc.) on job offers.

Group D variables reflect local Hispanic and black non-Hispanic population proportions. These population proportions are also multiplied by years of school or potential experience in order to capture possible interactions. Group E variables are interactions between human capital variables (schooling and potential experience) and minority status. Group F variables are triple-interaction effects, i.e., minority indicators multiplied both by minority population proportions and by either years of school or years of potential experience.

Since the CPDF is similar to the personnel data files of a single employer, the variable list for the regression analyses based on these data includes more detailed information on the individual's work history. The variable list does not include the detailed educational, language, and immigrant backgound data found in the SIE. The variables used in the regressions based on the CPDF are as follows:

Dependent Variables

natural logarithm of annualized salary

Independent Variables

Group A (race and ethnicity indicators):

1 if Hispanic, 0 otherwise

1 if black, 0 otherwise

Group B (expanded race and ethnicity indicators):

1 if Hispanic, 0 otherwise

1 if black, 0 otherwise

1 if Oriental, 0 otherwise

1 if American Indian, 0 otherwise

Group C (human capital, geographic location, etc.):

educational attainment indicators (1 if possesses the indicated characteristics, 0 otherwise) for each of the following mutually exclusive categories:

completed elementary school, did not complete high school

has some high school education, but did not complete high school

has high school diploma or equivalent

attended terminal occupational training program, but did not complete it

completed terminal occupational training program

attended less than one year of college

attended one year of college

attended two years of college

has associate-in-arts or equivalent degree

attended three years of college

attended four years of college, but did not receive B.A. or equivalent degree

has B.A. or equivalent degree

has B.A. or equivalent and some post-B.A. training

has first professional degree (e.g., J.D., M.D.)

has first professional degree and some post-first-professional-degree training

has M.A. or equivalent degree

has M.A. or equivalent and some post-M.A. training

has a sixth-year degree (e.g., Advanced Certificate in Education)

has a sixth-year degree and some post-sixth-year degree training

has Ph.D. or equivalent degree

has Ph.D. or equivalent degree and some post-Ph.D. training years since highest degree, for persons with at least a B.A. or equivalent (for persons with less than a B.A., this variable is set at zero)

square of years since highest degree

indicators for field of highest degree, for persons with at least a B.A. or equivalent (1 if field of highest degree is the one indicated and zero otherwise; set at zero for all persons with less than a B.A.), as follows:

medical doctors (M.D., D.D.S., D.V.M., etc.)

allied health professions (nursing, therapy, etc.)

mathematics, architecture, engineering, data processing

physical or biological sciences

arts or humanities

social sciences

law

age

square of age

years employed in federal government

square of years employed in federal government

product of age and years employed in federal government

1 if has physical or mental disability, 0 otherwise

indicators for veterans' preference (1 if possesses the indicated type of veterans' preference, 0 otherwise), as follows:

five-point veterans' employment preference

ten-point disability veterans' employment preference

ten-point compensable veterans' employment preference

ten-point other veterans' employment preference (e.g., spouse, survivor)

indicators for state of residence (1 if lives in a particular state, 0 otherwise) for all 48 states in the continental U.S. and the District of Columbia

The Group A and Group B variables differ only in that, in the latter group, we distinguish between Orientals and American Indians, on the one hand, and all other persons who are neither black nor Hispanic, on the other. More or less by definition, this group of all other persons might be called majority white.

Note that our Group C variables (reflecting human capital, geographic location, and the like) are quite similar to the ones used in our SIE regression models in some respects, but are rather different in other respects. In particular, the CPDF data permit us to derive educational attainment indicators that are more detailed than the ones that can be obtained from the SIE data: for example, the latter do not contain any measures of the number of years elapsed since highest degree, or of the field of the highest degree, while the CPDF data do; and while the SIE measures the number of years of school completed, the CPDF data provide somewhat more information about the amount and kind of educational attainment than the simple amount of time spent in school. The CPDF data also contain a measure of years of employment in the federal government, while the SIE data do not contain any measure of actual work experience, even with one's present employer. Of course, on the other hand, the CPDF data do not contain measures of some variables of interest that are available in the SIE. For example, the CPDF data do not contain any information on language skills and also do not differentiate between race or ethnicity. That is, the SIE data classify persons according to both race and ethnicity (which, for example, permits one to differentiate between black and white Hispanics), while in the CPDF classification scheme race and ethnicity are defined in such a way as to make black and Hispanic mutually exclusive.

We use the variables listed above to form two different regression models. The first model uses the simple Group A race-ethnicity indicators and the Group C variables, while the second model uses the expanded Group B race-ethnicity indicators and the Group C variables. Note that the first model, comprising Group A and Group C variables, is most comparable to the basic model used in our SIE regressions.

Because of the problems associated with estimating many parameters in logit models we use a smaller set of the available variables in our analysis of labor force status. The variable list for the logit analyses based on the SIE data is as follows:

Dependent Variable

labor force status, categorized as follows:

employed in the federal sector

employed in the non-federal sector

unemployed

not in the labor force

Independent Variables

number of years of formal education

potential experience (= age minus years of schooling minus 5)

1 if age is over 30 and under 41, 0 otherwise

1 if age is over 40 and under 51, 0 otherwise

1 if age is over 50, 0 otherwise

number of years lived in U.S. (equal to age, for persons born in U.S.)

1 if born in U.S., spoke English as a child, and speaks English now, 0 otherwise

1 if married with spouse present, 0 otherwise

number of persons in household

percent of population in area (classified by state, SMSA, and central city) that is Hispanic

percent of population in area that is black non-Hispanic

percent Hispanic in area times years of school

percent black non-Hispanic in area times years of school

percent Hispanic in area times potential experience

percent black non-Hispanic in area times potential experience

1 if of Puerto Rican origin, 0 otherwise

We estimate various logit models, containing alternative combinations of these variables, separately for each sex, using separate samples of Hispanics, black non-Hispanics, and white (that is, other) non-Hispanics. Note that an indicator for Puerto Rican ethnicity cannot be included in logits for samples of black or white non-Hispanics because, by definition, this indicator has a value of zero for all such persons. On the other hand, we do include such an indicator in logits for samples of Hispanics in order to distinguish between Puerto Ricans and other Hispanics.

LABOR FORCE STATUS RESULTS

One of our principal interests in this research is to compare the federal and non-federal sectors. The implications of our logit models with respect to employment in these two sectors are summarized in Table 1, which compares the actual and predicted employment sector for each

Table 1

Comparison of Minorities' Predicted Employment Proportions
(Predictions use logit coefficients from the sample of whites)

	Federal Employees			Private Employees		
	Actual %	Predicted %	% Diff.	Actual %	Predicted %	% Diff.
Men						
Hispanic	4.53	3.57	26.9	75.82	75.08	1.0
Puerto Rican	4.58	2.49	83.9	68.64	77.22	-11.1
Hispanic non-Puerto Rican	4.53	3.71	22.1	76.89	77.22	2.8
Black	5.07	2.98	70.1	67.62	73.17	-7.6
White	3.91	--	--	78.41	--	--
Women						
Hispanic	1.65	1.62	1.9	46.61	47.45	-1.8
Puerto Rican	1.35	1.00	35.0	34.28	51.64	-33.6
Hispanic non-Puerto Rican	1.69	1.71	-1.2	48.43	47.06	2.9
Black	3.53	1.29	173.6	50.96	52.08	-2.2
White	1.60	--	--	43.95	--	--

Note: Data base is the Survey of Income and Education; see text for description of analysis.

race/ethnic group. The comparison is based on the characteristics of each individual in the sample (regardless of actual sector). A predicted probability was generated using the estimated logit coefficients for labor force status from the white sample. All comparisons in this table concerning under- or overrepresentation are made relative to white non-Hispanics (men, in the case of the other male ethnic groups; or women, in the case of the other female ethnic groups). A positive entry for a given sector in the column headed "% Diff." indicates that the group in question is overrepresented in that sector relative to white non-Hispanics of the same sex with the same educational attainment, age, etc.; a negative entry indicates underrepresentation.

The main implications of Table 1 may be summarized as follows. First, virtually all minority ethnic groups (that is, groups other than white non-Hispanics) are substantially overrepresented in federal employment relative to white non-Hispanics. (The only exceptions to this generalization are Hispanic non-Puerto Rican women, who are slightly underrepresented in federal employment, and Hispanic women generally, who are only slightly overrepresented in federal employment, on average.) However, note that such overrepresentation in federal employment is only a small proportion of any given group's population. (For example, Table 1 indicates that men of Puerto Rican origin are overrepresented in federal employment in the sense that the actual proportion of such men in federal employment is 4.58%, as opposed to the 2.49% that would be expected if this group acted and were treated in regard to labor force status as white men with identical schooling, age, etc.) In this sense, an end to such overrepresentation would not involve the reallocation of a large number of persons. Second, Puerto Ricans of either sex are also substantially underrepresented in non-federal employment relative to comparable white non-Hispanics. Third, black non-Hispanic males are also underrepresented in non-federal employment. Recall that these differences in labor force status cannot be attributed exclusively to either supply or demand factors (e.g., to individual tastes or to employer discrimination) since the estimated version of the logit model does not identify either of these two behavioral relationships separately

To complement Table 1, we present in Table 2 a summary of the implications of our logit results concerning the relation between ethnicity and nonemployment, i.e., either unemployment or absence from the labor force. This shows that both men and women in each of the minority ethnic groups considered in our analyses are overrepresented among the unemployed, relative to whites with comparable schooling, age, family composition, etc. Non-Puerto Rican Hispanics of either sex and black women tend to be underrepresented among persons not in the labor force; Puerto Ricans of either sex, and black men, tend to be overrepresented.

All things considered, our logit results suggest that ethnicity as such does not have a particularly pronounced association with labor force status once one holds constant the effects of other supply and demand factors such as age, schooling, family composition, and the like. One simple way to illustrate this is shown in Table 3. In this table, we show how changing the ethnicity of <u>all</u> ethnic groups to white non-Hispanic (without changing their age, schooling, etc.) would alter the distribution of our total sample by labor force status. As shown there, changing the ethnicity of all persons in our sample to white would

Table 2

Comparison of Minorities' Predicted Unemployment and Not-in-Labor-Force Proportions
(Predictions use logit coefficients from the sample of whites)

	Unemployed			Not in Labor Force		
	Actual %	Predicted %	% Diff.	Actual %	Predicted %	% Diff.
Men						
Hispanic	5.83	5.43	7.4	18.03	20.23	-10.9
Puerto Rican	8.92	7.24	23.2	21.51	17.09	25.9
Hispanic non-Puerto Rican	5.43	5.20	.2	17.58	20.63	-14.8
Black	7.16	5.20	37.7	27.38	21.64	26.5
White	3.45	--	--	22.54	--	--
Women						
Hispanic	4.90	4.96	7.5	52.80	53.52	-1.3
Puerto Rican	6.37	4.87	30.8	62.74	49.73	26.2
Hispanic non-Puerto Rican	4.70	4.52	4.0	51.44	54.04	-4.8
Black	7.21	4.78	50.8	41.20	46.69	-11.8
White	3.32	--	--	51.13	--	--

Note: Data base is the Survey of Income and Education; see text for description of analysis.

Table 3

Comparison of Predicted and Actual Labor Force
Distribution for Entire Sample
(Predictions use logit coefficients from the sample of whites)

	Men (N = 10,025)		Women (N = 11,361)	
	Actual %	If White, Predicted %	Actual %	If White, Predicted %
Labor Force Status				
Employed	72.86	73.94	45.78	44.89
In federal sector	4.40	3.57	2.09	1.53
In non-federal sector	68.46	70.37	43.69	43.36
Unemployed	5.22	4.62	4.86	4.14
Not in labor force	21.92	21.44	49.36	50.97
Ethnicity				
Hispanic	38.51		37.91	
Puerto Rican	4.36		4.56	
Other	34.15		33.35	
Black	22.99		24.16	
White	38.49		37.93	

Note: Data base is the Survey of Income and Education; see text for description of analysis.

produce rather small shifts in the distribution of our total sample by labor force status. For example, the proportion unemployed among men would fall about 0.6 of a percentage point, while the proportion unemployed among women would fall by about 0.7 of a percentage point. (Recall, also, that our total sample for each sex consists of roughly equal numbers of Hispanics and white non-Hispanics, with somewhat smaller numbers of black non-Hispanics. Thus, minorities are substantially overrepresented in our sample relative to their representation in the population—meaning that any changes of the kind shown in Table 3 would be much smaller in the actual population than they are in our sample.)

DIRECT REGRESSION RESULTS FROM THE SIE DATA

In this section we discuss our direct (conventional least squares) regression results on ethnic pay differences, taking each sex in turn. (See J. Abowd and Killingsworth, 1981, for detailed tables; A. Abowd, 1982, for alternative specifications.)

Results for Men

1. Magnitudes. The pay differential for a given ethnic group relative to comparable white non-Hispanics varies considerably by ethnic group (and, as noted below, to a lesser extent by sector). All differentials are negative, implying that minority ethnic groups tend to be paid less than whites who are otherwise comparable (in terms of the other variables in the regression model from which the differential is derived). They are largest in absolute value (between about -.14 to

-.25) for black non-Hispanics, smallest in absolute value (between about -.01 to -.05) for non-Puerto Rican Hispanics, and of intermediate size (between about -.07 to -.13) for Puerto Ricans.

2. Statistical significance. As measured by t-statistics, the statistical significance of these pay differentials is generally quite substantial for blacks (t-ratios for most black-white differentials are between about 5.9 and 9.9); t-ratios for most Puerto Rican-white differentials are considerably lower (between about 0.8 and 2.2). Most differentials between non-Puerto Rican Hispanics and comparable whites would not be judged statistically different from zero at conventional levels (t-ratios for most of these differentials are between about 0.3 and 1.5).

3. Sectoral patterns. For all three minority ethnic groups, minority-white differentials in wages are larger in absolute value (that is, more negative) in the federal sector than in the non-federal sector, while minority-white differentials in earnings are larger in the non-federal sector than in the federal sector. For example, the black-white wage differential in the federal versus non-federal sector is about -.16 to -.18 (-.14), while the comparable figure for the earnings differential in the federal versus non-federal sector is about -.16 to -.17 (-.25).

4. Alternative dependent variables. For all three minority groups, the wage differential is larger than the earnings differential in the federal sector, but smaller than the earnings differential in the non-federal sector. For example, for Puerto Ricans, the wage (earnings) differential is about -.12 to -.13 (-.08 to -.10) in the federal sector, while in the non-federal sector the wage (earnings) differential is about -.07 to -.08 (-.13).

5. Alternative models. For all three minority groups, estimates of a given differential are relatively robust with respect to alternative models (that is, use of alternative sets of independent variables). For example, regression estimates of the Puerto Rican-white differential are about -.08 to -.12 when population proportion variables are not included, and are about -.10 to -.13 when such variables are included among these regressors. (Changes in differentials for most other race/ethnic groups attendant upon inclusion of these variables are smaller still.)

Results for Women

1. Magnitudes. With a few exceptions, minority-white pay differentials among women are smaller than among men, and many are either positive (implying that certain groups of minority women are paid more than comparable white women) or else essentially zero, in a statistical sense. The black-white pay differential among women is about .05 to -.05; the Puerto Rican-white female pay differential is about .12 to about -.40; and the non-Puerto Rican Hispanic-white pay differential is about .04 to -.13.

2. Statistical significance. On the whole, the statistical significance of minority-white pay differentials, as measured by their t-ratios, is lower among women than among men. Black-white differentials among women have t-ratios in the range 0.6 to 2.1; Puerto Rican-white differentials have t-ratios between [illegible] and 1.6; and non-Puerto Rican Hispanic-white differentials have t-ratios between .7 and 2.3.

3. Sectoral patterns. With a few exceptions, minority-white pay differentials among women are more negative (that is, lower in absolute

value) in the federal than in the non-federal sector. For example, the black-white pay differential is about -.04 to -.05 in the federal sector, while differentials in the non-federal sector are between about .05 and -.05.

4. Alternative dependent variables. Black-white and non-Puerto Rican Hispanic-white differentials in wages are typically more negative (that is, larger in absolute value if negative, or smaller in absolute value if positive) than are earnings differentials; while in the case of Puerto Rican-white differentials just the reverse holds. For example, the Puerto Rican-white wage differential is about .06 to .12, while the earnings differential is about -.01 to -.40; black-white and non-Puerto Rican Hispanic-white differentials in wages (earnings) are about -.02 to -.13 (.04 to -.09).

5. Alternative models. For all three minority groups, estimates of pay differentials are relatively robust with respect to alternative models (that is, use of alternative sets of independent variables). For example, regression estimates of the Puerto Rican-white wage (earnings) differential are about .06 to .12 (-.01 to -.40) when population proportion variables are not included, and about .06 to .14 (-.01 to -.39) when such variables are included among the regressors in a given model.

REVERSE REGRESSION RESULTS FROM THE SIE DATA

In this section we present the results of our reverse regression analysis for each sex. (See J. Abowd and Killingsworth, 1981, for detailed tables and A. Abowd, 1982, for a discussion of alternative specifications.)

Results for Men

1. Magnitudes. The pay differential for a given ethnic group relative to comparable white non-Hispanics varies considerably by ethnic group (and, as noted below, to a lesser extent by sector). Unlike the direct regression differentials, most of which are negative (implying that minorities tend to receive lower pay than comparable whites), most of the reverse regression differentials are positive (implying that minorities tend to receive higher pay than comparable whites). The black-white differential is between about .06 and -.05; the non-Puerto Rican Hispanic-white differential is between about .14 and .02; and the Puerto Rican-white differential is between about .06 and -.01.

2. Statistical significance. As measured by their t-statistics, the statistical significance of these pay differentials is generally quite substantial for non-Puerto Rican Hispanics (t-statistics for this group are between about 3.7 and 5.9). Black-white differentials in the federal sector, and Puerto Rican-white differentials in the non-federal sector, also have relatively high t-ratios (between about 3.4 and 4.2, and between about 1.5 and 3.1, respectively). However, black-white differentials in the non-federal sector and Puerto Rican-white differentials in the federal sector would not generally be judged different from zero, in a statistical sense, at conventional levels of significance.

3. Sectoral patterns. The magnitudes and even signs of these differentials vary considerably by sector. Puerto Rican-white differentials are always smaller in algebraic value (either negative, or else positive but small) in the federal sector than in the non-federal sector (the

range of federal sector differentials is about .02 to -.01, while the non-federal sector differential is about .06). On the other hand, differentials between non-Puerto Rican Hispanics and comparable whites in the non-federal sector (which are in the range .07 to .02) are smaller than the differentials in the federal sector (which are in the range .14 to .09). Finally, introducing population proportion variables changes completely the sectoral pattern of the black-white differentials. In models in which these variables are not included, the black-white differential in the federal versus non-federal sector is -.04 to -.05 (.02 to .01), but when such variables are included the differential in the federal versus non-federal sector is between about .04 and .06 (.01 and .00).

4. Alternative dependent variables. For all three minority groups, the wage differential is about the same as the earnings differential both in the federal and in the non-federal sector. For example, for Puerto Ricans, the wage (earnings) differential is about .02 to .01 (.00 to -.01) in the federal sector, while in the non-federal sector the wage and earnings differentials are both about .06 and .05.

5. Alternative models. For Puerto Ricans and other Hispanics, estimates of differentials are relatively robust with respect to alternative models (that is, use of alterative sets of independent variables). On the other hand, the federal black-white differential seems to be fairly sensitive to inclusion of population proportion variables. When such variables are excluded, the federal (non-federal) black-white pay differential is between about -.04 and -.05 (.02 and .01), and when such variables are included, the differential is between about .06 and .04 (.01 and .00).

Results for Women

1. Magnitudes. Minority-white pay differentials among women exhibit few obvious patterns; very roughly speaking, there appear to be about as many positive differentials (implying that minority women are paid more than comparable white women) as negative differentials (implying that minority women are paid less than comparable white women), and a large number do not appear to be different from zero (in a statistical sense) at conventional levels of significance. The black-white pay differential is between about .04 and -.17; the Puerto Rican-white pay differential is between about .14 and -.11; the non-Puerto Rican Hispanic-white pay differential is between about .08 and -.01.

2. Statistical significance. On the whole, the statistical significance of minority-white pay differentials, as measured by their t-ratios, is lower among women than among men. Black-white differentials among women have t-ratios in the range 1.1 to 8.6; Puerto Rican-white differentials have t-ratios between .4 and 5.1; and non-Puerto Rican Hispanic-white differentials have t-ratios between .2 and 7.6.

3. Sectoral patterns. With a few exceptions, minority-white pay differentials among women are lower in algebraic value (that is, larger in absolute value if negative, and smaller if positive) in the federal than in the non-federal sector. For example, the black-white pay differential is about -.02 to -.17 in the federal sector, while differentials in the non-federal sector are between about .04 and .02.

4. Alternative dependent variables. Black-white and Puerto Rican-white differentials in earnings are typically more negative (that is, larger in absolute value if negative, or smaller in absolute value if

positive) than are wage differentials. For example, the Puerto Rican-white wage differential is about .14 to .03, while the earnings differential is about -.04 to -.11. On the other hand, non-Puerto Rican Hispanic-white differentials in wages are greater in algebraic value in the federal sector, and are smaller in the non-federal sector, than are non-Puerto Rican Hispanic-white differentials in earnings.

5. <u>Alternative models</u>. For all three minority groups, estimates of pay differentials seem fairly robust with respect to alternative models (that is, use of alternative sets of independent variables). For example, regression estimates of the Puerto Rican-white wage (earnings) differential are about .14 to .03 (-.04 to -.11) when population proportion variables are not included, and about .13 to .05 (-.05 to -.08) when such variables are included among the regressors in a given model.

STRUCTURAL REGRESSION RESULTS FROM THE SIE DATA

We now discuss our structural (instrumental variable) regression results for each sex. (See J. Abowd and Killingsworth, 1981, for detailed tables, and A. Abowd, 1982, for a discussion of alternative specifications.)

<u>Results for Men</u>

1. <u>Magnitudes</u>. The pay differential for a given ethnic group relative to comparable white non-Hispanics varies considerably by ethnic group (and, as noted below, to a lesser extent by sector). In most cases, these differentials are negative (implying that minority groups

are paid less than comparable whites), and many of them are quite close to the corresponding direct wage regression differential. (We say more about this below.) Differentials are largest in absolute value (between about -.14 to -.25) for black non-Hispanics, smallest in absolute value (between about -.01 to -.05) for non-Puerto Rican Hispanics, and of intermediate size (between about -.07 to -.14) for Puerto Ricans.

2. Statistical significance. As measured by their t-statistics, the statistical significance of these pay differentials is generally quite substantial for blacks (t-ratios for most black-white differentials are between about 6.2 and 10.4); t-ratios for most Puerto Rican-white differentials are considerably lower (between about 1.0 and 2.1). Most differentials between non-Puerto Rican Hispanics and comparable whites would not be judged statistically different from zero at conventional levels (t-ratios for most of these differentials are between about 0.3 and 1.5).

3. Sectoral patterns. For all three minority ethnic groups, minority-white differentials in wages are larger in absolute value (that is, more negative) in the federal sector than in the non-federal sector, while minority-white differentials in earnings are larger in the non-federal sector than in the federal sector. For example, the black-white wage differential in the federal versus non-federal sector is about -.18 to -.20 (-.14), while the comparable figure for the earnings differential in the federal versus non-federal sector is about -.17 to -.19 (-.25).

4. Alternative dependent variables. For all three minority groups, the wage differential is larger than the earnings differential in the federal sector, but smaller than the earnings differential in the non-federal sector. For example, for Puerto Ricans, the wage (earnings)

differential is about -.14 (-.11) in the federal sector, while in the non-federal sector the wage (earnings) differential is about -.07 to -.08 (-.13).

5. Alternative models. For all three minority groups, estimates of a given differential are relatively robust with respect to alternative models (that is, use of alternative sets of independent variables). For example, regression estimates of the Puerto Rican-white differential are about -.08 to -.14 when population proportion variables are not included and are about -.07 to -.14 when such variables are included among the regressors.

Results for Women

1. Magnitudes. With a few exceptions, minority-white pay differentials among women are smaller than among men; most are fairly similar to the corresponding direct wage regression estimate; many are essentially zero, in a statistical sense. The black-white pay differential among women is about .06 to -.09; the Puerto Rican-white female pay differential is about .29 to -.53; and the non-Puerto Rican Hispanic-white pay differential is about .03 to -.36.

2. Statistical significance. On the whole, the statistical significance of minority-white pay differentials, as measured by their t-ratios, is lower among women than among men. Black-white differentials among women have t-ratios in the range 0.1 to 1.7; Puerto Rican-white differentials have t-ratios between 0.2 and 0.9; and non-Puerto Rican Hispanic-white differentials have t-ratios between 0.4 and 2.0.

3. Sectoral patterns. Black-white and non-Puerto Rican Hispanic-white pay differentials among women are usually somewhat more negative (that is, lower in absolute value) in the federal than in the non-federal sector. For example, the black-white differential is about -.00 to -.09 in the federal sector, while differentials in the non-federal sector are between .06 and -.02. Finally, the Puerto Rican-white differential is always larger in absolute value in the federal sector than it is in the non-federal sector--but the estimated wage differentials imply that Puerto Ricans are paid more than comparable whites, particularly in the federal sector, while the estimated earnings differentials imply that Puerto Ricans are paid less than comparable whites, especially in the federal sector. (See below.)

4. Alternative dependent variables. Black-white and non-Puerto Rican Hispanic-white differentials in wages are typically more negative (that is, larger in absolute value if negative, or smaller in absolute value if positive) than are earnings differentials; while in the case of Puerto Rican-white wage differentials just the reverse holds. For example, the Puerto Rican-white wage differential is about .25 to .29 in the federal sector (vs. about .05 in the non-federal sector), while the differential in earnings in the federal sector is about -.39 to -.53 (vs. -.02 to -.03 in the non-federal sector).

5. Alternative models. For all three minority groups, estimates of pay differentials are fairly robust with respect to alternative models (that is, use of alternative sets of independent variables). For example, regression estimates of the black-white wage (earnings) differential are about .05 to .29 (-.02 to -.53) when population proportion variables are not included and about .05 to .25 (-.03 to .39) when such variables are included among the regressors in a given model.

DIRECT AND REVERSE WAGE REGRESSION RESULTS FROM THE CPDF DATA

In this section we discuss direct and reverse regression results derived from the federal government CPDF data.

1. Results by race/ethnicity and sex. In general, the CPDF results seem fairly similar to the SIE results as regards racial and ethnic pay differentials by sex within the federal sector. As in the SIE results, the CPDF results imply that both Hispanics and blacks are paid less within the federal sector than are whites (that is, either non-black non-Hispanics, including American Indians and Orientals as well as majority whites; or majority whites as such). In general, black-white pay differentials in the CPDF results are larger in absolute value than Hispanic-white pay differentials; and, for either racial-ethnic group, the minority-white differential among men is larger than the minority-white differential among women. Most of the CPDF differentials are statistically different from zero at reasonable levels of significance.

2. Results by type of statistical model. In our CPDF results, as in our SIE results, reverse wage regression generally produces estimates of differentials that are less negative than those derived using direct wage regression; indeed, in several instances (notably for Hispanics), the direct wage regression estimate of the minority-white differential has a negative sign (implying that minority persons are paid less than comparable whites), but the reverse wage regression estimate is positive (implying that minority persons are paid more than comparable whites). Black women are an exception to this generalization, however; in some cases, the reverse wage regression estimate of the black-white differen-

tial for women is slightly more negative than the corresponding direct wage regression estimate. Finally, the shrinkage in the estimated differential (that is, the extent to which use of reverse wage regression makes a given differential less negative) seems, in general, to be smaller in the CPDF data than in the SIE data.

3. Comparison with results derived from the SIE. On the whole, both the direct and reverse wage regression estimates of the black-white differential derived from the CPDF are similar to the direct and reverse wage regression estimates of this differential derived from the SIE. (However, the CPDF direct wage regression black-white differentials among men seem somewhat smaller, in absolute value, than the corresponding SIE estimates.) On the other hand, the CPDF estimates of the Hispanic-white differential seem, in general, to be somewhat closer to zero (either smaller if positive, or less negative, if negative) than the corresponding SIE estimates. However, the differences between the SIE and CPDF estimates do not, in general, seem particularly large.

COMPARISON OF ALTERNATIVE ESTIMATORS AND RESULTS

We now consider the alternative estimation techniques that we have used in evaluating the determinants of pay. We do this using our preferred results from the SIE for men and for women, which we set out in Tables 4 and 5, respectively. These results are all evaluated at the mean values of all variables for white non-Hispanics and are derived from either our basic regression model (in which case they are labeled "without population proportions") or from our detailed regression model with population proportions but without three-way interactions (in which case they are labeled "with population proportions").

In drawing conclusions about our three different estimation techniques from Tables 4 and 5, it is worth recalling that these techniques are concerned with different statistical and conceptual issues. First, structural regression is concerned with estimating the answer to the first methodological question; that is, with estimating differences in employer wage offers. It does not, however, make a correction for possible measurement error bias. Second, both direct and reverse wage regressions are concerned with estimating the answer to the second methodological question; that is, with estimating differences in compensation conditional on employment. Direct regression does not make a correction for possible measurement error bias while reverse regression does make a correction of this kind. Thus, it would be reasonable to expect that these three different techniques would produce different results. The key issue is, of course, the extent to which results derived from these techniques do in fact differ.

As before, it seems advisable to consider each sex separately. As regards men, it is evident from Table 4 that the reverse regression differentials contrast sharply with both the direct and the structural regression differentials: differentials estimated using either of the latter two techniques are usually negative (and often significantly different from zero, in a statistical sense), while differentials estimated using the former technique are frequently positive. As regards the federal sector, both structural and direct regression differentials are negative, but the latter are usually somewhat smaller in absolute value

Table 4

Summary of Direct, Reverse, and Structural Wage Regression Estimates of Ethnic Differentials in Pay for Men Evaluated at Mean Values of Whites

	Federal Sector			Non-Federal Sector		
	Hispanics of Puerto Rican Origin	Hispanic non-Puerto Ricans	Blacks	Hispanics of Puerto Rican Origin	Hispanic non-Puerto Ricans	Blacks
A. Without Population Proportion Variables						
1. Log Wages						
direct	-.1241	-.0466	-.1789	-.0799	-.0265	-.1426
	(.0820)	(.0321)	(.0183)	(.0481)	(.0232)	(.0192)
reverse	.0201	.1097	-.0471	.0619	.0509	.0215
	(.0497)	(.0186)	(.0111)	(.0201)	(.0085)	(.0897)
structural	-.1413	-.0513	-.1987	-.0783	-.0257	-.1409
	(.0856)	(.0342)	(.0191)	(.0482)	(.0232)	(.0191)
2. Log Earnings						
direct	-.0857	-.0186	-.1650	-.1340	-.0441	-.2476
	(.1072)	(.0419)	(.0238)	(.0628)	(.0303)	(.2503)
reverse	.0042	.1362	-.0517	.0479	.0403	.0133
	(.0682)	(.0255)	(.0151)	(.0323)	(.0137)	(.0145)
structural	-.1083	-.0404	-.1887	-.1344	-.0437	-.2472
	(.1118)	(.0447)	(.0250)	(.0629)	(.0303)	(.0250)
B. With Population Proportion Variables						
1. Log Wages						
direct	-.1279	-.0476	-.1612	-.0697	-.0080	-.1420
	(.0821)	(.0337)	(.0204)	(.0482)	(.0241)	(.0202)
reverse	.0056	.0945	.0390	.0818	.0682	.0093
	(.0521)	(.0195)	(.0116)	(.0211)	(.0090)	(.0095)
structural	-.1444	-.0503	-.1750	-.0676	-.0070	-.1386
	(.0856)	(.0356)	(.0214)	(.0482)	(.0241)	(.0202)
2. Log Earnings						
direct	-.0965	-.0108	-.1564	-.1304	-.0443	-.2497
	(.1073)	(.0440)	(.0267)	(.0630)	(.0315)	(.0264)
reverse	-.0140	.1083	.0584	.0587	.0235	.0030
	(.0705)	(.0264)	(.0156)	(.0325)	(.0137)	(.0146)
structural	-.1134	-.0278	-.1726	-.1299	-.0438	-.2475
	(.1118)	(.0465)	(.0280)	(.0631)	(.0315)	(.0264)

Note: Data base is the Survey of Income and Education; standard errors are in parentheses.

Table 5

Summary of Direct, Reverse, and Structural Wage Regression Estimates of Ethnic Differentials in Pay for Women Evaluated at Mean Values of Whites

	Federal Sector			Non-Federal Sector		
	Hispanics of Puerto Rican Origin	Hispanic non-Puerto Ricans	Blacks	Hispanics of Puerto Rican Origin	Hispanic non-Puerto Ricans	Blacks
A. Without Population Proportions						
1. Log Wages						
direct	.1209 (.1576)	-.1268 (.0567)	-.0501 (.0236)	-.0590 (.0619)	-.0267 (.0262)	-.0119 (.0189)
reverse	.0320 (.0765)	.0127 (.0264)	-.0512 (.0110)	.1398 (.0276)	.0659 (.0183)	.0329 (.0095)
structural	.2943 (.3268)	-.3615 (.1784)	-.0025 (.0387)	.0472 (.0621)	-.0290 (.0263)	-.0147 (.0189)
2. Log Earnings						
direct	-.3962 (.2522)	-.0327 (.0907)	-.0378 (.0377)	-.0075 (.1074)	.0350 (.0454)	.0537 (.0327)
reverse	-.1133 (.1340)	.0574 (.0462)	-.1665 (.0194)	-.0394 (.0258)	.0176 (.0196)	.0444 (.0182)
structural	-.5281 (.5908)	.1889 (.3224)	-.0945 (.0699)	-.0185 (.1077)	.0332 (.0456)	.0563 (.0329)
B. With Population Proportions						
1. Log Wages						
direct	.1363 (.1575)	-.1328 (.0586)	-.0485 (.0263)	.0622 (.0621)	-.0199 (.0273)	-.0149 (.0195)
reverse	.0548 (.0820)	-.0050 (.0283)	-.0205 (.0118)	.1349 (.0281)	.0802 (.0105)	.0174 (.0097)
structural	.2474 (.3264)	-.3516 (.1759)	-.0045 (.0381)	.0508 (.0622)	-.0242 (.0273)	-.0173 (.0195)
2. Log Earnings						
direct	-.3997 (.2523)	-.0911 (.0938)	-.0507 (.0421)	-.0144 (.1077)	.0350 (.0473)	.0402 (.0338)
reverse	-.0754 (.1392)	.0208 (.0480)	-.1228 (.0202)	-.0513 (.0548)	.0581 (.0203)	.0216 (.0189)
structural	-.3938 (.5011)	-.1211 (.3025)	-.0112 (.0655)	-.0256 (.1080)	.0321 (.0474)	.0415 (.0339)

Note: Data base is the Survey of Income and Education; standard errors are in parentheses.

than the former. On the other hand, structural and direct regression differentials for the non-federal sector, while usually negative, are also generally quite close to each other; indeed, in many instances, a structural wage regression differential for the non-federal sector is usually slightly smaller than its direct wage regression counterpart, although the difference is generally very small. Finally, in most instances (particularly as regards the federal sector), t-ratios for structural wage regression differentials are somewhat larger than t-ratios for their direct wage regression counterparts: standard errors of estimated structural wage regression differentials are slightly larger than standard errors of estimated direct wage regression differentials, but the estimates themselves are larger still, particularly for the federal sector.

While Table 4 thus suggests a variety of generalizations concerning the impact of using alternative estimation techniques as far as estimates for men are concerned, Table 5, for women, suggests little in the way of patterns or stylized facts. The three estimation techniques, applied to the federal sector, seem to produce three rather different sets of estimated ethnic differentials among women. Estimates for the non-federal sector derived using the three techniques seem, on the whole and roughly speaking, to be somewhat closer together. However, in many cases—and to a much greater extent than is true of our results for men—the differentials for women reported in Table 5 would not be judged different from zero, at conventional levels of significance, regardless of the technique used in estimating them. In this sense, then, the results of these different estimation techniques are closer together than cursory inspection of Table 5 might suggest.

Table 6 compares the results obtained from both the SIE and the CPDF for the year 1975. For the two estimation techniques considered, direct and reverse, the results from these data sources are quite similar. Essentially the same inferences are supported in either data set.

SUMMARY AND CONCLUSIONS

There is not much consistent or compelling evidence in our results to suggest that minority women generally suffer substantial wage discrimination (in either the Question 1 or Question 2 sense) relative to comparable white women. One possible exception to this statement concerns black women in the federal sector, where our results usually show negative pay differentials. (However, a considerable number of these differentials do not differ from zero, in a statistical sense, at reasonable levels of significance.) An important caveat in this respect is that our data do not contain measures of actual work experience (Garvey and Reimers, 1980). We are, therefore, forced to use a proxy, potential experience.

Second, as regards ethnic differentials in pay among men, our results suggest (a) that minority men may suffer discrimination both in terms of conditional differentials and in terms of offers, and (b) that estimates of the magnitudes of both kinds of discrimination may be subject to serious measurement error bias. Part (a) of this conclusion follows in a straightforward way from consideration of our direct and structural wage

Table 6

Comparison of Ethnic Pay Differentials for Men and Women for 1975 Derived from SIE and CPDF Data

	Men		Women	
	SIE	CPDF	SIE	CPDF
Hispanics				
direct	-.0558	-.0543	-.1020	-.0134
	(.0304)	(.0080)	(.0542)	(.0114)
reverse	.0993	.0283	.0146	-.0017
	(.0176)	(.0062)	(.0251)	(.0107)
Blacks				
direct	-.1787	-.1381	-.0503	-.0603
	(.0182)	(.0130)	(.0236)	(.0151)
reverse	-.0471	-.0421	-.0512	-.0441
	(.0111)	(.0110)	(.0110)	(.0147)

Notes: Standard errors are in parentheses. SIE columns present regression differentials derived from the Survey of Income and Education for men and women in the federal sector; dependent variable = natural logarithm of hourly wages. CPDF columns present regression differentials derived from the federal Central Personnel Data File; dependent variable = natural logarithm of annualized salary.

regression results; note that our results provide much stronger support (in the sense of statistical significance) for this proposition with respect to blacks than with respect to Puerto Rican or other Hispanics. Part (b) of this conclusion is prompted by our reverse wage regression results.

Third, our results also suggest that wage discrimination against minority males (particularly blacks) is greater in the federal than in the non-federal sector, while earnings discrimination against minority males (particularly blacks) is smaller in the federal than in the non-federal sector. At first sight, this may seem paradoxical: if the non-federal sector is better than the federal sector as regards wage discrimination, why isn't it also better as regards earnings discrimination? One possible explanation of this apparent paradox has to do with employment instability, which is greater in the non-federal sector than in the federal sector: if minorities suffer substantially and disproportionately (relative to comparable whites) from the relatively greater employment instability (layoffs, etc.) in the non-federal sector, then the non-federal sector could well be worse than the federal sector as regards earnings differentials even if it is better as regards wages. Our logit results on labor force status appear to suggest that minority groups generally are overrepresented among the unemployed. While this finding does not prove the validity of our conjecture about sectoral patterns in wage vs. earnings differentials, it is certainly consistent with it.

Of course, the notion that discrimination within the federal sector may be substantial is not new. Our results not only support this view but also suggest something else: discrimination against minority males, particularly in terms of wages and with respect to blacks, is of <u>greater</u> magnitude in the federal than in the non-federal sector. This is particularly noteworthy because previous studies have tended to suggest just the opposite. We suspect that one reason for this is that, in contrast with previous work, we have attempted to control in a fairly detailed fashion for purely geographic effects on pay (via differences in the cost of living and the like). Since minorities are generally overrepresented in federal employment, and since such federal employment is concentrated in urban areas in particular states, sorting out purely geographic effects on pay (in effect, purely compensating or equalizing premia) from other kinds of effects, including ethnicity, obviously need not be a trivial matter. Indeed, the difference between our results and those found in previous work suggests that such effects may be important.

NOTES

[1]Studies that attempt to decompose earnings differentials into portions attributable to employer discrimination and portions attributable to differences in productivity characteristics such as education include, among others, Blinder (1973), Oaxaca (1973), and Smith (19 , Litigation under Title VII of the Civil Rights Act and other antidiscrimination laws and regulations is implicitly or explicitly concerned with the extent to which observed employment and earnings differences between sexes or between racial or ethnic groups are attributable to employer discrimination per se rather than to other factors such as differences in productivity-related characteristics. Analyses of earnings differences in the context of legal proceedings include Baldus and Cole (1980), Ehrenberg (1979), and Finkelstein (1980).

[2]One important reason for studying employment and earnings differences by sector is that such differences may reveal the extent to which a particular sector is unusual compared to the rest of the economy. (For example, see Smith, 1977.) A second reason is that nonpecuniary rewards to employment may vary by sector: for example, federal government employment may entail greater job security or better working conditions than employment elsewhere in the economy (Smith, 1977). We define wage discrimination as a differential in the total reward to employment, including both pecuniary and nonpecuniary rewards. This reinforces the usefulness of an intrasectoral analysis of wage discrimination since important differences in nonpecuniary compensation across sectors are, in effect, held constant. On the other hand, the fact that such an analysis

may have conceptual advantages over an intersectoral study does not necessarily mean that statistical procedures suitable for the latter kind of study are also suitable for the former kind of study.

REFERENCES

Abowd, Anthony M. 1982. Minority pay differentials in the federal and non-federal sectors. Unpublished manuscript, Graduate School of Business, University of Chicago.

Abowd, John M., and Mark R. Killingsworth. 1981. An analysis of Hispanic employment, earnings and wages in the federal and non-federal sectors with special reference to Puerto Ricans. Technical Report no. 21-36-78-61, Department of Labor, Employment and Training Administration. Washington, D.C.: GPO.

Arrow, Kenneth J. 1973. The theory of discrimination. In Orley C. Ashenfelter and Albert Rees, eds., Discrimination in labor markets. Princeton, N.J.: Princeton University Press.

Bald··s, David C., and James W. L. Cole. 1980. Statistical proof of discrimination. New York: McGraw-Hill.

Blinder, Alan S. 1973. Wage discrimination: Reduced form and structural estimates. Journal of Human Resources, 12, 550-560.

Borjas, George J. 1978. Discrimination in HEW: Is the doctor sick or are the patients healthy? Journal of Law and Economics, 21, 97-110.

Chiswick, Barry R. 1978. The effect of Americanization on the earnings of foreign-born men. Journal of Political Economy, 86, 897-922.

Chiswick, Barry R. 1980. The earnings of white and coloured male immigrants in Britain. Economica, 47, 81-87.

Ehrenberg, Ronald G. 1979. The regulatory process and labor earnings. New York: Academic Press.

Finkelstein, Michael O. 198?. The judicial reception of multiple regression studies in race and sex discrimination cases. Columbia Law Review, 80, 737-754.

Garvey, Nancy, and Cordelia Reimers. 1980. Predicted vs. actual work experience in an earnings function for young women. In Ronald G. Ehrenberg, ed., Research in labor economics, Vol. 3. Greenwich, Conn.: JAI Press.

Hashimoto, Masanori, and Levis Kochin. 1979. A bias in the statistical estimation of the effects of discrimination. Economic Inquiry, 18, 478-486.

Heckman, James J. 1979. Sample selection bias as a specification error. Econometrica, 47, 153-162.

Heckman, James J., Mark R. Killingsworth, and Thomas MaCurdy. 1981. Empirical evidence on static labour supply models: A survey of recent developments. In Zmira Hornstein, Joseph Grice, and Alfred Webb, eds., The economics of the labour market. London: Her Majesty's Stationery Office.

Kamalich, Richard F., and Solomon W. Polachek. 1982. Discrimination: Fact or fiction? An examination using an alternative methodology. Unpublished manuscript, Department of Economics, University of North Carolina.

Kiefer, Nicholas M., and Sharon P. Smith. 1977. Union impact and wage discrimination by region. Journal of Human Resources, 12, 521-534.

Kmenta, Jan. 1971. Elements of econometrics. New York: Macmillan.

Maddala, G. S. 1977. Econometrics. New York: McGraw-Hill.

Malkiel, Burton G., and Judith A. Malkiel. 1973. Male-female pay differentials in professional employment. American Economic Review, 63, 693-705.

McFadden, Daniel. 1973. Conditional logit analysis of qualitative choice behavior. In Paul Zarembka, ed., Frontiers in econometrics. New York: Academic Press.

McFadden, Daniel. 1975. The revealed preferences of a government bureaucracy: Theory. Bell Journal of Economics, 6, 401-416.

Mincer, Jacob. 1974. Schooling, experience and earnings. New York: Columbia University Press, for the National Bureau of Economic Research.

Oaxaca, Ronald. 1973. Male-female wage differentials in urban labor markets. International Economic Review, 14, 693-709.

Oaxaca, Ronald. 1976. Male-female wage differentials in the telephone industry. In Phyllis Wallace, ed., Equal employment opportunity and the AT&T case. Cambridge, Mass.: MIT Press.

Osterman, Paul. 1979. Sex discrimination in professional employment: A case study. Industrial and Labor Relations Review, 32, 451-464.

Roberts, Harry V. 1979. Harris Trust and Savings Bank: An analysis of employee compensation. Report 7946, Center for Mathematical Studies in Business and Economics, University of Chicago.

Roberts, Harry V. 1980. Statistical biases in the measurement of employment discrimination. In E. Robert Livernash, ed., Comparable worth: Issues and alternatives. Washington, D.C.: Equal Employment Advisory Council.

Roberts, Harry V. 1981. Reverse regression and employment discrimination: Rationale and criticisms. Unpublished manuscript, Graduate School of Business, University of Chicago.

Schneider, Philip A.D. 1974. The central personnel data file. Unpublished manuscript, Bureau of Management Information Systems, U.S. Office of Personnel Management.

Smith, Sharon P. 1977. Equal pay in the public sector: Fact or fantasy? Princeton, N.J.: Industrial Relations Section, Princeton University.

U.S. Bureau of the Census. 1978. Microdata from the survey of income and education. Data Access Description No. 42, Bureau of the Census, U.S. Department of Commerce.

Welch, Finis. 1973. Black-white differences in returns to schooling. American Economic Review, 63, 893-907.